

Métodos de Previsão

Por

João Pedro Fernandes Pais

Dissertação de Mestrado em Economia e Administração de Empresas

Orientada por:

Professora Doutora Alexandra Patrícia Horta Ramos

Professor Doutor José Fernando Gonçalves

2017

Nota biográfica

João Pedro Fernandes Pais nasceu em Viseu em 1990. Terminou em 2013 a licenciatura Economia na Universidade de Coimbra. Atualmente, frequenta o Mestrado em Economia e Administração de Empresas na Faculdade de Economia da Universidade do Porto desde setembro de 2013, com o intuito de complementar a sua formação académica.

A nível profissional iniciou funções na MCoutinho Peças em setembro de 2015 como gestor de marca/stock.

Agradecimentos

Em primeiro lugar, gostaria de agradecer aos meus orientadores, Professora Alexandra Ramos e ao Professor José Gonçalves, que sempre se mostraram disponíveis para ajudar no desenvolvimento da presente dissertação. Um especial agradecimento pela revisão cuidadosa de todos os documentos que elaborei e pelos comentários e sugestões que permitiram o enriquecimento do conteúdo da presente dissertação.

Mencionar a contribuição do António Fernandes e do Hélder Paiva por dar a conhecer os métodos de previsão, nos dados disponibilizados e do incentivo dado no desenrolar desta dissertação.

Deixo ainda um grande obrigado à Natália, por todas as sugestões, paciência, apoio e carinho.

Um agradecimento especial à Marta e à Suzi pela batalha constante, pelo apoio e pelas noites passadas nas pós-graduações.

Finalmente, quero sublinhar o contributo dos meus pais, dos meus avós, da minha irmã e dos meus amigos, pelo apoio, paciência e motivação demonstrados.

Resumo

A previsão constitui o ponto de partida para o processo de planeamento de toda a atividade futura, tomada de decisão e controlo em qualquer organização.

A previsão dos valores futuros de uma série temporal, designada por previsões multi-passos, reveste-se de muita importância uma vez que permite o planeamento antecipado e correto dos sistemas em estudo. A previsão em séries temporais tem sido influenciada principalmente pela econometria e estatística. Nas últimas décadas, a aprendizagem automática, *machine learning*, uma área dedicada ao desenvolvimento de algoritmos que aprendem a partir dos dados, tem-se mostrado muito ativa em modelação e predição.

A crescente procura por valores mais altos no nível de serviço e de qualidade superior pressiona as organizações a usufruírem de maior flexibilidade e a direcionarem as suas atividades para a satisfação do consumidor, com a finalidade de melhorar a velocidade e a confiança da resposta, delineando os seus procedimentos e sistemas para esse objetivo.

Nesta dissertação são analisados os consumos das séries relativas à procura de peças na MCozinha Peças com o objetivo de averiguar qual dos métodos de previsão presentes na literatura deverá ser utilizado e decidir qual deles terá melhor performance de entre os seguintes métodos: Método ARIMA, Métodos de alisamento exponencial, Método de Holt-Winters e o Método de Croston.

Palavras-chave: Stock, Métodos de previsão; ARIMA; ETS; Holt-Winters; Croston; RStudio.

Abstract

Forecasting is the starting point for the planning process of all future activity, decision making and control in any organization. The uncertainty associated with multi-step forecast allows accurate knowledge of their behaviour in the future, which leads us to make better predictions.

Forecasting future observations from a time series is a problem of increasing importance. Time series forecasting has been influenced mainly by econometrics and statistics. In the last decades, machine learning, an area dedicated to the development of algorithms that learn from data, has become one of the most active areas of predictive modelling research.

The increasing demand for higher values in the service level and higher quality pressures organizations to enjoy greater flexibility and to direct their activities towards consumer satisfaction, to improve response speed and confidence, outlining the procedures and systems for this purpose.

In this dissertation, will be analysed the consumption of the series related to the demand for parts in M Coutinho Peças with the objective of ascertaining which the forecasting methods will be used and deciding which one will perform best among the following methods: ARIMA Method, Exponential smoothing methods, Method of Holt-Winters and the Croston Method.

Keywords: Stock; Forecast; ARIMA; ETS; Holt-Winters; Croston; RStudio.

Índice

Nota biográfica.....	i
Agradecimentos	ii
Resumo.....	iii
Abstract	iv
Índice de figuras.....	vii
Índice de tabelas.....	viii
Acrónimos.....	ix
Tabela de Símbolos Matemáticos	x
1. Introdução	11
1.1. Enquadramento	11
1.2. Objetivos do trabalho	12
1.3. Estrutura do Documento	13
2. Revisão da literatura.....	14
2.1. Gestão de stocks: definição, objetivos e custos.....	14
2.1.1 Custo de stocks.....	15
2.1.2 Sistemas de controlo de stock	16
2.1.3 Classificação de stocks.....	22
2.1.3 Compromisso entre o custo e detenção	23
2.2. Métodos de Previsão	25
2.2.1 Modelos Causais	25
2.2.2 Série temporal	26
2.3 Tópicos básicos de previsão.....	28
2.3.1 Estacionariedade	28
2.3.2 Função Autocorrelação	30
2.3.3 Função Autocorrelação Parcial	31
3. Metodologia	33
3.1 Processos Autorregressivos.....	33
3.2 Média Móvel.....	34
3.3 Processos Autorregressivos e de Médias Móveis	35
3.4 Processo Autorregressivo Integrado de Média Móvel	36

3.4.1 SARIMA	38
3.5 Métodos de alisamento exponencial	39
3.5.1 Métodos de alisamento exponencial simples	39
3.5.2 Método de Croston	40
3.5.3 Método de Holt	41
3.5.4 Método de Holt-Winters	42
3.6 Erros de previsão.....	44
3.7 Seleção do modelo	45
4. Caso de estudo	47
4.1 Descrição da amostra	49
4.2. Modelação e análise quantitativa	49
4.3 Recolha e preparação dos dados	50
4.4 Tratamento dos dados	51
4.4.1 Classificação dos artigos por classe	51
4.4.2 Avaliação da estacionariedade	53
4.4.3 Implementação da gestão de stocks	54
5. Análise dos resultados.....	56
5.1 Análise descritiva das séries.....	56
5.2 Comparação entre os erros dos vários métodos	58
5.3 Teste para comparação das médias dos erros dos diferentes métodos	62
5.4 Exemplos da aplicação dos métodos.....	65
6. Conclusões	67
7. Referências bibliográficas	69
8. Anexos	73

Índice de figuras

Figura 1 - Revisão contínua de stocks. (Fonte: Waters, 2003)	19
Figura 2 - Revisão periódica de stocks. (Fonte: Waters, 2003)	20
Figura 3 - Compromisso entre o custo de detenção de stock e o custo de indisponibilidade na estimação do nível ótimo do stock. (Cavalieri et al. 2008)	23
Figura 4 - Série com correlação de curta duração	31
Figura 5 - Série com dependência de longa duração.....	31
Figura 6 - Série com sazonalidade	31
Figura 7 – Diagrama Box-Jenkins.....	46
Figura 8 – Organograma Grupo MCoutinho.....	48
Figura 9 – Ambiente de modelação e previsão em RStudio	50
Figura 10 – Análise ABC de 7635 referências com consumo no ano de 2016.....	52
Figura 11 - Série temporal “Filtro Óleo”	56
Figura 12 - Decomposição da série temporal “Filtro óleo”	57
Figura 13 - Gráfico ACF da série original “filtro de óleo”	57
Figura 14 - Erros MSE das 90 referências em estudo	59
Figura 15 - Erros MAE das 90 referências em estudo	59
Figura 16 - Erros MAPE das 90 referências em estudo	60
Figura 17 – Série intermitente, espigão de plástico – modelo ARIMA (0,1,1).....	65
Figura 18 – Série intermitente, espigão de plástico – Método de Croston.....	65
Figura 19 – Amortecedor – modelo ARIMA (0,1,1)	66
Figura 20 - Amortecedor – Método de Croston	66

Índice de tabelas

Tabela 1 – Comparação entre os vários Sistema de controlo de stock (Fonte: própria)	21
Tabela 2 – Erro quadrático médio (MSE)	60
Tabela 3 – Erro médio absoluto (MAE).....	61
Tabela 4 – Percentagem Erro Médio Absoluto (MAPE)	61

Acrónimos

ACF	Autocorrelation Function
ADF	Augmented Dickley-Fuller
AIC	Akaike's Information Criterion
AICc	Corrected Akaike's Information Criterion
AR	Autoregressive model
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive–moving-average model
BIC	Bayesian Information Criterion
CR	Croston
EOQ	Economic Order Quantity
ETS	Exponential Smoothing
HW	Holt-Winters
IID	Independentes e identicamente distribuídas (variáveis)
KPSS	Teste Kwiatkowski-Phillips-Schmidt-Shin
MA	Moving-average model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
PP	Teste Phillips-Perron
SKU	Stock Keep Unit

Tabela de Símbolos Matemáticos

Notação Matemática	Descrição
ε_t	Erro associado à previsão no instante t
S_t	Componente Sazonal no instante t
T_t	Componente tendência-ciclo no instante t
Y_t	Valor observado no instante t
\hat{Y}_t	Previsão da série temporal para o instante t
\hat{Y}_{t+k}	Previsão corrigida para a série k no instante t
\bar{Y}	Valor médio da série temporal

1. Introdução

1.1. Enquadramento

Nas empresas modernas atuais, a gestão de *stocks* desempenha um papel primordial, revelando-se uma das temáticas cruciais pelo seu impacto na tesouraria e na saúde financeira geral da empresa, seja pelas suas consequências no nível de serviço¹, como também na satisfação proporcionada aos clientes. Consequentemente, se o desempenho das funções de gestão de existências for bem-sucedido, é possível fornecer níveis apropriados de serviço ao cliente, sem custos e sem níveis de *stock* exagerados.

Através da gestão de *stocks*, o que se pretende atingir é o nível de *stock* mais baixo que permita o nível de serviço mais alto (dualidade). Para alcançar este objetivo devemos investir em tecnologia, reorganizar os procedimentos e diminuir os custos associados. Assim sendo, a função da gestão de *stocks* consubstancia-se na definição dos produtos a encomendar, na identificação do tempo em que devem ser encomendados e na respetiva quantidade, devendo ser adotado o modelo de gestão de *stocks* que for mais vantajoso.

A incerteza, associada a muitos fenómenos temporais, não permite o conhecimento exato do seu comportamento no futuro, o que nos leva a fazer previsões. Qualquer plano de negócio depende da previsão de vendas. Os gestores de marketing e vendas têm de elaborar regularmente, planos previsionais de vendas de cada um dos produtos ou serviços da empresa, tendo em conta as flutuações da procura e da oferta. Para isso, deve valer-se da sua experiência, consultar opiniões qualificadas e recorrer a métodos estatísticos de modo a evitar desperdícios de recursos e potenciais custos de oportunidade.

O gestor de compras e produção, responsável pelo planeamento e organização de todo o processo de aquisição de mercadorias, matérias-primas e equipamentos necessários à produção e comercialização, vai estar dependente das vendas dos níveis de cada, SKU, *stock keeping unit* (artigos, linhas de produto, pacotes, serviços, taxas, etc.), por categoria de produto ou por cliente.

¹ Nível de serviço entende-se como o número de unidades que podem ser fornecidas no momento a partir do *stock* disponível. (Chase *et al.*, 2006)

1.2. Objetivos do trabalho

Esta dissertação tem por base um caso de estudo na *MCoutinho Peças*, empresa que comercializa peças originais de automóveis. O negócio de *peças originais*, que era até então gerido pelas concessões de cada marca, passa assim a ter uma nova estrutura com gestão centralizada dentro do *Grupo MCoutinho*. Os benefícios deste modelo foram desde logo evidentes, destacando-se: orientação para o cliente, gestão centralizada de *stocks*, portefólio muito alargado de marcas, modelo de distribuição eficiente e atendimento telefónico especializado.

A empresa *MCoutinho Peças* enfrenta vários desafios relacionados com a gestão de *stocks*. O seu desafio principal prende-se com a distinção de empresas concorrentes pela capacidade de resposta, assumindo papel de *first choice* dos seus clientes. Posto isto, este tipo de posicionamento perante os clientes torna mais urgente a gestão de *stocks*.

Como contributo e considerando que a *MCoutinho Peças* quer sempre satisfazer os seus clientes com a maior celeridade, revela-se premente analisar o problema da gestão de *stocks*. Ao definir uma estratégia de otimização do *stock* e da sua alocação pode proporcionar uma melhoria da eficiência financeira. Ciente das lacunas na gestão de *stocks*, têm sido discutidas algumas medidas que poderão possibilitar uma melhor otimização. Apesar dessa preocupação não ser recente, é agora mais importante, devido ao aumento da exigência dos mercados e da necessidade de eficiência financeira.

Os objetivos desta dissertação levantaram várias questões de investigação que se pretendem ver respondidas no final deste trabalho, designadamente: Como se pode prever a procura nos próximos períodos? Será que a previsão da procura obtida vai de encontro com a realidade?

Atualmente, dado o aumento da procura, revela-se cada vez mais difícil dar resposta atempada aos clientes e manter, no mesmo período de tempo, um nível de *stock* sustentável. Presentemente, as empresas operam num ambiente extremamente competitivo onde os padrões da procura são cada vez mais irregulares e esporádicos. E a venda de *peças originais* não é exceção à regra, pois também é caracterizada por padrões de procura intermitente. Estes padrões de procura são caracterizados pela procura dispersa em determinados intervalos de tempo e apresentam dificuldades consideráveis em termos

de previsão e controlo de *stock*, tendo em conta não ser conhecido quando é que a próxima procura vai ocorrer e qual será o seu respetivo volume (Babai *et al*, 2011).

A complexidade em lidar com estes tipos de padrões de procura está em encontrar o melhor *trade-off* entre os efeitos negativos relacionados com a manutenção de níveis de armazenamento elevados, nomeadamente: necessidade de um armazém com dimensões superiores e os recursos para manter essa dimensão; os elevados custos de detenção destes níveis de *stock*; os riscos de obsolescência de determinados artigos e os efeitos negativos relacionados com níveis de armazenamento baixos, que levam à perda de clientes e da procura (Gamberini *et al*, 2010).

1.3. Estrutura do Documento

A estrutura da presente dissertação encontra-se dividida por vários capítulos. No primeiro capítulo engloba a presente introdução, indicando a motivação para o problema, e os objetivos.

No segundo capítulo é feita a revisão da literatura e são apresentados os principais objetivos da previsão e a importância de prever numa organização empresarial.

De seguida o terceiro capítulo, metodologia, serão abordados alguns dos métodos de previsão presentes na literatura e as suas diferentes etapas de previsão.

No quarto capítulo apresenta-se o tratamento e caracterização do conjunto de dados abordado neste trabalho e a análise ABC dos dados obtidos.

O quinto capítulo corresponde à análise dos resultados alcançados e as respetivas conclusões no sentido de se averiguar e descobrir os melhores métodos de previsão.

Finalmente apresentarei no último capítulo, as respetivas conclusões e trabalhos futuros.

2. Revisão da literatura

2.1. Gestão de *stocks*: definição, objetivos e custos

O conceito de *stocks* pode ser definido como a “existência de qualquer artigo usado numa organização”. Os *stocks* são mantidos para criar segurança face a atrasos na entrega de matérias ou produtos por parte dos fornecedores, para aumentar a segurança perante grandes variações na procura e obter vantagem da dimensão económica de uma ordem e compra. (Chase *et al.*, 2006)

O excesso desta existência de artigos, *stock*, acaba por ter um impacto negativo nos custos e com isso diminuir o poder competitivo da empresa. Estes *stocks* devem então ser controlados por um conjunto de políticas e controlos que fiscalizem os níveis de *stocks* a serem mantidos, bem como quando e em que quantidade se deve encomendar. (Chase *et al.*, 2006)

Segundo Gonçalves (2010), existem ainda diversos tipos de *stock*, consoante a natureza do negócio da organização: i) *stocks* necessários ao fabrico que inclui matérias-primas, matérias subsidiárias, embalagens e materiais de embalagem; ii) *stocks* de conservação, os quais respeitam às peças sobresselentes necessárias para as máquinas, ferramentas e matérias consumíveis; iii) *stocks* em curso de fabrico, que correspondem a produtos não concluídos e que já implicaram o consumo de recursos; iv) *stocks* sazonais² constituídos para enfrentar picos de procura sazonais ou ruturas na capacidade produtiva e, finalmente, v) *stocks* de segurança, *stocks* criados para fazer face a situações de incerteza.

O conceito de *stock* de segurança está relacionado com a determinação do momento ideal em que uma nova encomenda ou ordem de produção são inseridas. Tal facto surge devido à procura e o prazo de entrega serem variáveis inconstantes, o que gera algum grau de incerteza. Para reduzir essa incerteza, são determinados níveis de *stocks* de segurança, recorrendo-se, geralmente, a previsões para estimar o nível da procura.

² Existência de variações da procura acima do normal, levando a que por vezes seja vantajoso produzir para *stock*, em momentos cuja procura é baixa, para depois conseguir satisfazer a procura em alturas de forte aumento, em que a produção não é suficiente, recorrendo-se ao *stock* excedentário.

Contudo, as previsões podem falhar e converterem-se em dois tipos de imprevisto, designadamente, consumo previsional inferior ao real, o que implica que ocorram ruturas e consumo previsional superior ao real, aumentando os custos de posse. As ruturas de *stock* para a generalidade das empresas, podem ser mais problemáticas do que a existência de *stock* em excesso. Por conseguinte, muitas vezes, as empresas optam por estabelecer *stocks* mínimos ou de segurança.

Uma rutura de stock apresenta normalmente as seguintes consequências:

- Perda de venda (custo de oportunidade) ou paragem de fabricação, onde o custo é possível de calcular (custos tangíveis);
- Perda de imagem junto do mercado, cujo custo é extremamente difícil de calcular (custos intangíveis).

A presença de um stock de segurança implica um aumento do custo de posse. Este custo adicional pode, no entanto, ser compensado pela economia gerada pela não-ocorrência de ruturas.

Há, pois, que encontrar o seu justo valor, resultante de uma análise de compromisso. A qualidade de serviço, ou satisfação do cliente, é designada por “nível de serviço”. O risco de rutura representa a probabilidade de, quando solicitado, não ser possível satisfazer uma encomenda. (Gonçalves, 2010).

2.1.1 Custo de *stocks*

Como referem os autores Chase *et al.* (2006) quando se toma uma decisão que afete o nível de *stocks*, é necessário ter em consideração os custos que lhe são inerentes e que geralmente se classificam em:

- (i) custos de aprovisionamento;
- (ii) custos de posse;
- (iii) custos de rutura;
- (iv) custos de informação.

Custos de aprovisionamento são custos associados à aquisição de bens (encomenda) e determinam fortemente as quantidades económicas de reabastecimento dos inventários. Normalmente estes custos compreendem remunerações e encargos com os agentes,

estudos de mercado, despesas com negociações, controlo de prazos, transporte do produto, controlo das entregas e conferência das faturas.

Custos de posse são os custos associados à manutenção do inventário, estando intimamente ligados à posse dos artigos e à necessidade de armazenagem e, tipicamente, são diretamente proporcionais à dimensão média do inventário.

Custo de rutura ocorre sempre que uma encomenda de um cliente não pode ser satisfeita através das existências em *stock*.

E por último, os custos de informação associados à obtenção de informação para a tomada de decisões, os custos de utilização de um sistema informático e os custos ao realizar previsões da procura.

Portanto, conclui-se que, existem muitos fatores que favorecem a existência de inventários, no entanto e confrontando com os custos dos mesmos, é possível afirmar que é necessária uma gestão competente, para que se verifiquem os seguintes objetivos: uma resposta rápida e económica ao cliente. Para alcançar a eficiência da gestão de inventários, nas organizações, é necessário focar em duas questões fulcrais, entre elas: “Como controlar os stocks?” e “Que quantidades encomendar?” (Gonçalves, 2010). É com base nestas questões que serão introduzidos os modelos estocásticos de suporte à gestão de stocks.

2.1.2 Sistemas de controlo de *stock*

A variabilidade da procura é algo que se tem vindo a verificar como vulgar, para uma grande percentagem das empresas. Devido a fatores económicos, tecnológicos e estratégicos, a procura apresenta, frequentemente, comportamentos aleatórios, tanto a curto como a longo prazo.

Dada esta variação e a dificuldade de prever eficazmente o futuro, são utilizados modelos estocásticos, modelos que auxiliam o gestor a estudar e compreender o comportamento da procura, e aumentar a sua capacidade de resposta à imprevisibilidade da mesma, evitando ruturas de stock. (Gonçalves, 2010).

Os modelos estocásticos aplicam-se quando a procura e/ou oferta têm um comportamento aleatório, incerto. Esta incerteza aumenta a complexidade da gestão de

stocks, pois agora é necessário lidar com a possibilidade de existir rutura de stocks. Para lidar com este comportamento aleatório, tanto do lado da procura como do lado da oferta, é necessário construir um stock de segurança para absorver variações superiores aos valores registados. No entanto, estando a lidar com variáveis aleatórias, as variações que estas irão sofrer são imprevisíveis, o que significa que o stock de segurança consegue absorver algumas dessas variações, mas não necessariamente a totalidade. Quanto maior for o stock de segurança, maior é a probabilidade de ele conseguir absorver as variações imprevisíveis; no entanto, existirá sempre a probabilidade de ele não conseguir. Quando se fala em variáveis aleatórias, inevitavelmente terá de se falar em probabilidades. Neste contexto, o conceito de nível de serviço tem uma enorme importância para dimensionar o stock de segurança a constituir. Assim, Francischini *et al.* (2002), definem o stock de segurança, S_s , segundo a seguinte equação:

$$S_s = Z_x * D_p * \sqrt{P + P_e} \quad (0.1)$$

Onde:

Z_x constante normal padrão (tabelada em função do risco de rutura - R_r)

D_p desvio padrão da procura

P prazo médio de aprovisionamento (na mesma unidade de tempo do desvio-padrão da procura)

P_e periodicidade da encomenda

O nível de serviço é expresso em percentagem e corresponde à probabilidade de a empresa ter disponível a quantidade procurada, no momento procurado. Por exemplo, se a probabilidade de rutura de um artigo é 5%, o nível de serviço será de 95%. Portanto, o stock de segurança a constituir depende do nível de serviço (ou da probabilidade de rutura) que a empresa definir. Quanto maior for o nível de serviço que a empresa quer prestar aos seus clientes, maior será o stock de segurança a constituir.

O stock de segurança a constituir vai depender do modelo de gestão de stocks implementado. Nos modelos estocásticos existem dois modelos base:

- Modelo de revisão contínua, (Figura 1)
- Modelo de revisão periódica, (Figura 2)

Modelo de revisão contínua

O modelo de revisão contínua, baseia-se numa inspeção contínua da quantidade disponível para cada produto (Gonçalves, 2010), ou seja, esta ferramenta, constantemente, analisa os níveis de *stock*. Assim que estes stocks ficam abaixo de um nível pré-determinado, geralmente designado por ponto de encomenda ou nível de reabastecimento, a ordem de reposição é colocada.

Este sistema, averigua continuamente a quantidade disponível dos materiais e encomenda uma quantidade fixa (Q), assim que esta atinja um determinado nível de *stock* – ponto de encomenda (Pe). A data de encomenda é variável: quando a procura aumenta, o ponto de encomenda é atingido mais cedo; quando a procura diminui, o ponto de encomenda é atingido mais tarde. O ponto de encomenda (Pe) é definido como:

$$P_e = K * d + s \quad (0.2)$$

Onde:

K consumo mensal previsto por unidade de tempo

d prazo de entrega

s stock de segurança

A quantidade a encomendar, Q_e , procura encontrar um compromisso entre custos originados pela posse de stock e custos originados pela necessidade de aprovisionar³ e é definida por:

$$Q_e = \sqrt{\frac{2EN}{H}} \quad (0.3)$$

E preço de realização de encomenda

N consumo anual

H custo anual de posse e armazenamento

³ Aprovisionar, isto é, preparar, expedir e rececionar encomendas.

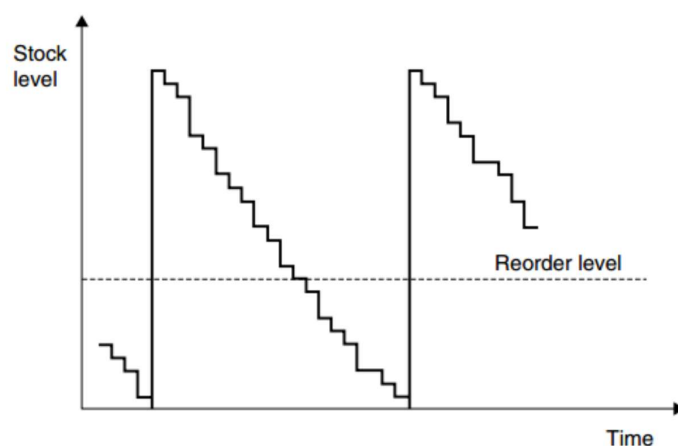


Figura 1 - Revisão contínua de stocks. (Fonte: Waters, 2003)

Na figura 1 é apresentado um exemplo de revisão contínua de stocks onde o nível de stock é dado em função do tempo e onde a linha a tracejado representa o ponto de encomenda.

As vantagens do método são desde logo o seu automatismo pois quando o nível de stock atinge uma quantidade pré-definida (ponto de encomenda) é sugerido o lançamento de uma encomenda para o fornecedor. As desvantagens do método é a sua difícil utilização quando surgem variações de consumo, pois terá que se alterar constantemente o ponto de encomenda. Não permite o fácil agrupamento de artigos na mesma encomenda, porque só artigos que são encomendados ao mesmo fornecedor atingem ponto de encomenda na mesma data. E se a encomenda não for lançada no momento em que o nível de stock atinge o ponto de encomenda, então o risco de rutura aumenta.

Modelo de revisão periódica

Neste modelo, também denominado método do ciclo de revisão constante, o momento de colocação da encomenda ao fornecedor é pré-definido (por negociação com o fornecedor ou por decisão interna), com uma periodicidade entre as encomendas fixa (semanal, quinzenal, mensal, entre outras). O período de aprovisionamento (PE) é assim constante (é igual à periodicidade económica) variando apenas a quantidade a encomendar de cada vez (Q_n), que se calcula segundo a seguinte equação:

$$Q_n = K(P + d) + s - (A + G) \quad (2.4)$$

Onde,

Q_n quantidade a encomendar

P consumo mensal previsto por unidade de tempo, periodicidade

d prazo de entrega

s stock de segurança

A stock existente em armazém

G quantidade encomendada ainda por entregar

A periodicidade da encomenda é uma constante e a sua fórmula é dada por:

$$PE = \frac{L_e}{\left(\frac{N}{12}\right)} \quad (2.5)$$

No dia estipulado para a colocação da encomenda, compara-se o stock existente e o stock necessário para o próximo período (stock alvo); a quantidade a encomendar corresponderá à diferença entre estes dois valores. O modelo designa-se por periódico pois os níveis de stock são revistos periodicamente e não continuamente.

Tal como os métodos de revisão contínua, os métodos de revisão periódica concentram-se na análise da quantidade de material existente, diferindo apenas no modo de como é realizada a inspeção.

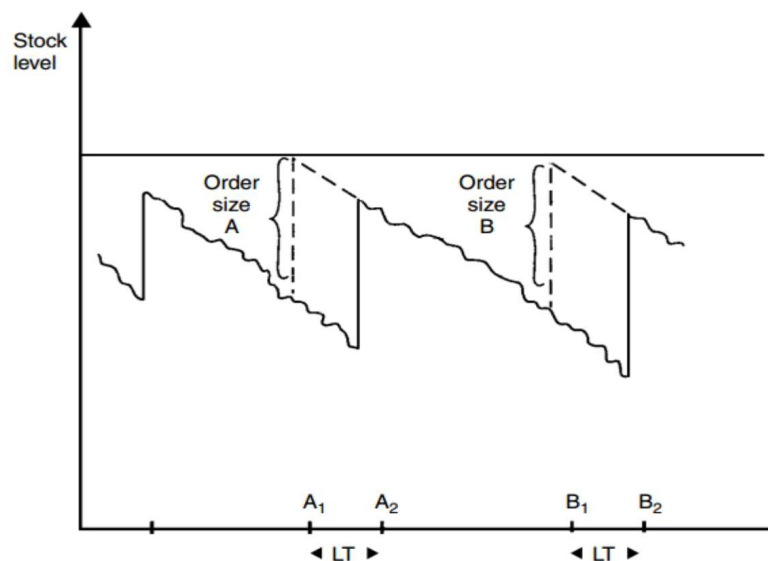


Figura 2 - Revisão periódica de stocks. (Fonte: Waters, 2003)

Na figura 2 é apresentado um exemplo de revisão periódica de stocks onde o nível de stock é dado em função do tempo e onde a linha a tracejado representa o dia estipulado para a encomenda.

As vantagens do modelo de revisão periódica começam na possibilidade de agrupar as encomendas de vários artigos de um mesmo fornecedor, possibilitando assim o estabelecimento de planos de aprovisionamento; permite uma distribuição igual das encomendas ao longo do ano, com a consequente vantagem de reduzir o custo de realização de encomenda e a possibilidade de descontos por quantidade e menores encargos de transporte. Possibilita também o ajuste às reais necessidades de consumo da empresa, pelas diferentes quantidades que são encomendadas de cada vez e que dependem da previsão do consumo e da existência em armazém.

Uma das desvantagens do modelo deve-se à sensibilidade a variações imprevistas e de prazo de entrega, que se forem relevantes podem originar ruturas de stock bem como o facto de ele se manter cego durante a periodicidade de aprovisionamento e, portanto, vulnerável às variações de consumo.

A literatura diz-nos que o modelo de revisão periódica é adequado quando se verifica pelo menos uma das seguintes condições:

- a) Variação muito pequena da procura e do prazo de aprovisionamento;
- b) Preço de custo de aquisição baixo;
- c) Interesse em encomendar em grupo a um mesmo fornecedor.

Tabela 1 – Comparação entre os vários Sistema de controlo de stock (Fonte: própria)

Sistemas de contro de stock	Quanto encomendar	Quando encomendar
Revisão contínua	Contante (L_e)	Variável
Revisão periódica	Variável	Constante ⁴

⁴ Constante pois é igual ao período económico.

2.1.3 Classificação de stocks

A gestão de stocks envolve anualmente milhares de movimentos de entradas e saídas de materiais. Para que possamos concentrar a nossa atenção de forma seletiva, torna-se necessário que o sistema de controlo de stocks e quebras proporcione uma forma de classificação que distinga os artigos mais importantes, em valor, dos menos importantes.

Portanto a elevada variedade de artigos e do número de unidades de cada um dos artigos existentes nos armazéns obriga-nos a prestar mais cuidada gestão aos que se revelam de maior importância em valor financeiro investido.

A análise ABC ou Análise de Pareto, desenvolvida com base nos estudos de Wilfred Pareto, define-se como um método que, em função de determinado critério permite pôr em evidência os elementos de uma população estatística aos quais se deve dedicar maior atenção por serem os mais relevantes. Pode também ser visto como um método de classificação de inventário, pois tem se revelado como uma ferramenta de gestão de grande valor na identificação dos produtos de stock com mais importância (Gonçalves, 2010).

A análise ABC permite identificar os produtos que terão um impacto significativo sobre o custo geral do inventário, ao mesmo tempo que fornece um mecanismo para identificar e classificar diferentes categorias de ações que exigirão um gerenciamento e controlo diferente. A classificação é feita em três grupos, de acordo com o grau de importância de cada produto, A, B e C. A categoria de cada produto é determinada a partir da sua contribuição para o valor do consumo anual, da seguinte forma:

- **Produtos A** – produtos que contribuem com uma maior percentagem do valor do consumo anual, mas representam uma pequena fração da totalidade dos produtos existentes. Tipicamente entre 15 a 25% dos produtos representam 70 a 80% dos custos anuais do valor financeiro investido em stocks.
- **Produtos B** – constitui um grupo intermédio, normalmente entre 20 a 25% do total de artigos representam aproximadamente 15% do valor.
- **Produtos C** – constitui o grupo de artigos que representa a maioria, mas que por outro lado fornecem uma pobre contribuição para o valor do consumo anual. Normalmente, 60 a 65% do total de artigos representam aproximadamente 5% dos custos anuais.

A classe A, contendo a maior soma de investimento, cerca de 70% a 80% e embora com o menor número de artigos, deve merecer maior atenção e vigilância. É necessário controlar a movimentação dos artigos deste grupo em referência.

A classe B, é um grupo intermédio exigindo uma menor vigilância. Contudo os artigos próximos da fronteira com a classe A devem ser vigiados dadas as variações que podem ocorrer no seu valor financeiro durante o ano de vigência da análise ABC e serem considerados artigos de classe A.

Classe C, não tem grande importância na gestão dos stocks dado o seu insignificante valor financeiro nos muitos artigos que constituem.

2.1.3 Compromisso entre o custo e detenção

Em resumo, a figura 3 ilustra que a disponibilidade do artigo na quantidade e momento certo é uma das preocupações centrais da logística, com o objetivo de assegurar ao cliente um serviço de nível satisfatório. A constituição de *stocks*, com frequência pode ser justificada pelas seguintes razões: apoio ao negócio em ambiente previsível, aproveitamento de oportunidades sazonais e proteção contra a incerteza. Todavia, a *performance* da empresa não deve ser posta em causa pela posse de um volume de *stocks* excessivo. (Fernie e Sparks, 2004; Nenes *et al.*, 2010; Fernandes *et al.*, 2013).

A quantidade de encomenda económica (*EOQ*), representada pela linha a tracejado cinzento, é assim a intersecção entre a recta do custo de posse (*Inventory holding costs*) e o custo de encomenda (*Ordering costs*) até ao custo total de aprovisionamento (*Total cost*).

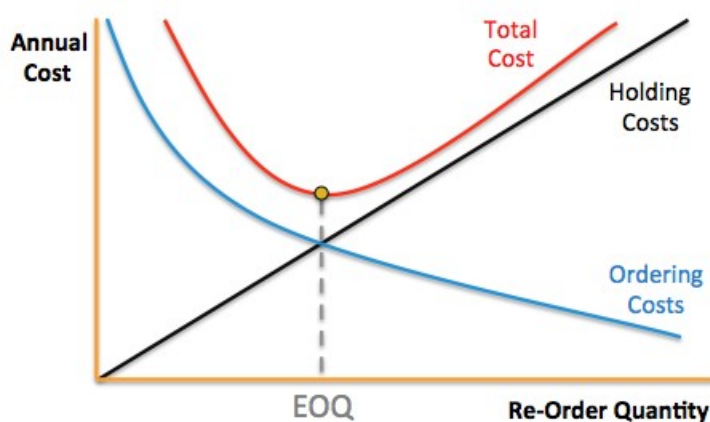


Figura 3 - Compromisso entre o custo de detenção de stock e o custo de indisponibilidade na estimação do nível ótimo do stock. (Cavaliere *et al.* 2008)

Face ao exposto, a gestão de *stocks* deve procurar equilibrar os dois lados da balança, visando, por um lado, garantir a satisfação do cliente através da disponibilidade do produto e, por outro lado, acautelar a situação financeira da empresa com uma adequada gestão dos custos logísticos, considerados excessivamente elevados por autores como (Fernie e Sparks, 2004) e Fernandes *et al.* (2013).

Martins *et al.* (2005) definem a previsão da procura como um processo metodológico para a determinação de dados futuros, baseado em modelos estatísticos, matemáticos ou econométricos.

O objetivo da análise de *stocks* é definir quando devem ser encomendados os artigos e qual a dimensão da encomenda. Para isso, é necessário prever a procura destes artigos de forma a garantir que estes estão disponíveis no momento em que serão necessários, distinguindo-se diferentes tipos de procura: (i) independente, (ii) dependente e (iii) intermitente, ou seja:

- (i) Estamos perante uma procura independente quando a procura de vários artigos não está inter-relacionada, devendo neste caso as quantidades serem determinadas separadamente. (Chase *et al.*, 2006).
- (ii) Se a necessidade de um qualquer artigo depende da existência de outro, estamos perante a procura dependente. (Chase *et al.*, 2006).
- (iii) A procura intermitente é caracterizada por intervalos de procura dispersos misturados com intervalos de tempo durante os quais não existe procura. O controlo do *stock*, face a estes padrões de procura, apresenta dificuldades consideráveis em termos de previsão. (M. Zied Babai, 2011). Exemplo dessa dificuldade é a sazonalidade associada a este tipo de artigos, o motivo pelo qual são usados modelos de previsão que contemplam a componente de sazonalidade em geral e normalmente associado ao tipo de dados que pretendemos estudar.

No planeamento e controlo de *stocks*, a incerteza da previsão da procura é um dos maiores problemas, pelo que os métodos de previsão são uma ferramenta fundamental para diminuir esta incerteza. (Gonçalves, 2010).

2.2. Métodos de Previsão

Face ao exposto, a indústria de peças automóveis tem procurado gerir a incerteza quanto aos níveis de previsão de procura, conduzindo ao desenvolvimento de vários métodos de previsão. Métodos estatísticos clássicos, como o alisamento exponencial e a regressão linear, têm vindo a ser usados durante várias décadas para prever estes níveis de previsão. (Bartezzaghi *et al.*, 1999)

Uma previsão consiste na tentativa de extrapolar o comportamento futuro a partir das condições atuais, utilizando dados históricos do passado. (Caiado, 2016)

A gestão de *stocks* recorre fundamentalmente a métodos quantitativos baseados em séries temporais. (Ballou, 2004)

Os Métodos de Previsão, ao preverem os níveis de procura, providenciam os *inputs* para o planeamento e controlo de stocks e podem ser um instrumento de auxílio à determinação dos níveis de *stock* de uma organização. (Ballou, 2004)

Os modelos de previsão segundo Gonçalves (2010) dividem-se em dois tipos:

- (i) Os modelos qualitativos, por se basearem em estimativas e opiniões, são considerados modelos subjetivos. Estes modelos são utilizados sempre que não existam dados históricos, ou caso existam, não sejam representativos do futuro. As técnicas qualitativas são utilizadas para previsões a médio e a longo prazo.
- (ii) Os modelos de previsão quantitativos assentam no conceito da utilização dos dados históricos da procura para estimar a procura no futuro. Urge identificar padrões nos dados históricos e transpor este padrão para o futuro. Dentro dos modelos quantitativos existem os modelos causais e os modelos de séries temporais.

2.2.1 Modelos Causais

A premissa na qual os modelos causais de previsão são construídos é a que a procura é descrita como um fator ou fatores subjacentes ao meio ambiente, isto é, o nível de previsão de uma variável deriva do nível da previsão de outras variáveis relacionadas com a primeira. Se, por exemplo, o nível de atendimento ao cliente tem um efeito positivo nas

vendas, então sabendo o nível de previsão do atendimento ao cliente, o valor de vendas pode ser projetado.

Os modelos causais podem ser estáticos no caso das regressões lineares e descritivos em modelos de simulação computacional.

No entanto, as variáveis causais são difíceis de estimar e encontrar e quando estas variáveis são encontradas, a sua previsão tende a ser muito baixa.

É então aconselhada a utilização dos métodos qualitativos em conjunto com os métodos quantitativos, para se proceder de forma mais eficaz, a uma análise crítica dos resultados e previsões obtidos. (Ballou, 2004)

2.2.2 Série temporal

Define-se uma série temporal como um conjunto de observações associadas a determinado fenómeno aleatório, efetuadas em períodos de tempo sucessivos e estatisticamente relacionadas. (Brockwell e Davis, 1991)

A análise de séries temporais é uma abordagem sistemática que permite não só compreender como uma variável se altera ao longo desse mesmo período de tempo, proporcionando assim uma descrição dos dados, como também permite estabelecer uma relação entre duas ou mais variáveis, ao longo de um determinado período de tempo.

A análise supracitada considera quatro componentes, a saber: tendência, sazonalidade, movimentos oscilatórios ou cíclicos e movimentos aleatórios (ruído).

2.2.2.1 Modelos de séries temporais

Ao analisar uma série temporal pretende-se, genericamente, alcançar dois objetivos, sendo eles, a modelação e a previsão dos valores futuros, os quais se encontram fortemente relacionados. O primeiro objetivo consiste em encontrar um modelo que tenha em conta as relações existentes entre as observações, permitindo a descrição da série temporal. O segundo objetivo diz respeito à previsão de valores futuros, pois como observa Murteira (2000) *"É a própria existência do tempo que leva as pessoas a fazer previsões e a tentar antecipar a evolução no futuro das sucessões que estudam"*.

As séries temporais, como refere Murteira (2000), podem ser decompostas e classificadas nas quatro componentes seguintes:

Tendência (T_t): a marcha principal da série, variação em média ao longo do tempo, descrevendo movimentos que se manifestam suavemente e consistentemente durante períodos longos. Esta componente descreve o comportamento mais notório da série (crescente, decrescente ou constante) durante o período de tempo considerado, isto é, define como um comportamento padrão de crescimento ou de redução, o qual se verifica durante um longo período de tempo.

Efeito Sazonal (S_t): variações ao longo da tendência que ocorrem com periodicidade curta, por exemplo 1 ano, isto é, são variações que se repetem consecutivamente, no mesmo período temporal. Está relacionada com as oscilações que ocorrem por período de tempo (diariamente, semanalmente ou mensalmente) no decorrer de um ano, podendo ser explicadas por diversos fatores como: as estações do ano, devido por exemplo, às diferenças de temperatura e ao turismo; medidas administrativas, justificando-se com o início ou final do mês; tradições; religião; festas, como por exemplo o Natal ou a Páscoa ou simplesmente devido a efeitos de calendário.

Efeito Cíclico (C_t): variações com caráter periódico, mas de período diferente do sazonal. Relacionam-se sobretudo com as fases de expansão e recessão dos sistemas económicos. Movimentos oscilatórios ou cíclicos que correspondem a ciclos como o económico, de expansão ou de recessão, constituindo assim, flutuações positivas e negativas não obedecendo a um determinado padrão temporal.

Efeito Ruído (ε_t): efeito aleatório, residual. Esta componente representa tudo o que não pode ser atribuído às outras três componentes.

Embora todas as séries temporais contenham a componente residual, as outras três componentes, podem não estar presentes na sua totalidade, por refletirem um determinado padrão comportamental que pode não ser explicado por qualquer um dos comportamentos acima explanados.

Face ao exposto, é possível fazer a reconstrução da série. A série pode assim ser escrita como função destas componentes, ou seja:

$$Y_t = f(T_t, S_t, C_t, \varepsilon_t) \quad (2.6)$$

As fórmulas matemáticas mais comuns podem ser da forma aditiva (Equação 2.7) ou multiplicativa (Equação 2.8):

$$Y_t = T_t + C_t + S_t + \varepsilon_t \quad (2.7)$$

$$Y_t = T_t * C_t * S_t * \varepsilon_t \quad (2.8)$$

Onde Y_t representa o valor observado no instante t , T_t representa a componente tendência, C_t efeito cíclico, S_t a componente da sazonalidade e ε_t o erro associado à previsão.

Enquanto o modelo aditivo se aplica normalmente às situações em que a amplitude dos valores de sazonalidade se mantém constante ao longo do tempo, o modelo multiplicativo é o mais adequado quando a série temporal apresenta variação na amplitude da sazonalidade.

Em suma, os objetivos da análise de séries temporais passam por descrever os dados, ajustando um modelo, construir novas teorias económicas, testar teorias económicas e prever valores futuros de variáveis económicas importantes.

De seguida descrevem-se os conceitos que serão usados posteriormente.

2.3 Tópicos básicos de previsão

2.3.1 Estacionariedade

Entenda-se por um processo estocástico $\{\varepsilon_t\}$ uma sequência de variáveis aleatórias indexadas no tempo. Matematicamente, um processo estocástico pode ser definido como uma coleção de variáveis aleatórias ordenadas no tempo e definidas em um conjunto de índices T , que pode ser contínuo ou discreto. (Brockwell e Davis, 1991)

Processo Estritamente Estacionário: O processo Y_t diz-se estritamente estacionário se para todo o n e todo t_1, \dots, t_n , a distribuição conjunta de Y_{t_1}, \dots, Y_{t_n} é igual à distribuição conjunta de $Y_{t_1+k}, \dots, Y_{t_n+k}$, qualquer que seja $k \in \mathbb{N}$.

Estacionariedade estrita implica que a distribuição de probabilidade conjunta seja constante ao longo do tempo, ou seja, que não dependa do instante t .

Processo Fracamente Estacionário ou Estacionário de Segunda Ordem: implica que os dois primeiros momentos (ou seja, a média e a matriz de variância-covariância) da

distribuição de probabilidade conjunta da variável estocástica sejam constantes ao longo do tempo, isto é: Processo Y_t com variância finita tal que:

- $E(Y_t) = \mu, \mu \text{ constante}$
- $V(Y_t) = \sigma^2 = cov(Y_t, Y_t)$
- $cov(Y_t, Y_s) = cov(Y_{t+r}, Y_{s+r}), \forall r, s, t$

Um processo estritamente estacionário é também estacionário de segunda ordem, mas o contrário nem sempre se verifica. Daqui em diante, entende-se por processo estacionário, o processo que for fracamente estacionário.

Processo não estacionário é um processo cuja média e/ou variância variam com o tempo. O facto desses dois momentos serem funções do tempo dificulta a possibilidade de fazermos inferência fora da amostra. Caso a estacionariedade não se verifique, poderá ser usada a diferenciação, que consiste em desfazar a série temporal em determinado número de instantes com o objetivo de estabilizar a média da série.

A estacionariedade é observada quando os dados não apresentam tendência crescente ou decrescente, nem movimentos periódicos, e flutuam em torno de uma média constante, independente do tempo, e a variância das flutuações não se altera ao longo do tempo. Contudo, uma série que não é estacionária em média também não é necessariamente estacionária em variância e covariância. (Wei, 2007)

Um processo estocástico $\{\varepsilon_t\}$ é chamado de ruído branco se for uma sequência de variáveis aleatórias não correlacionadas provenientes da mesma distribuição com média e variância constantes. (Murteira, 2000)

- $E(\varepsilon_t) = \mu_\varepsilon$ (usualmente $E(\varepsilon_t) = 0$)
- $V(\varepsilon_t) = \sigma_\varepsilon^2$
- $\varphi_k = cov(\varepsilon_t; \varepsilon_{t+k}) = 0$ para todo o $k \neq 0$
- $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$
 - Significa que ε_t são variáveis aleatórias independentes e identicamente distribuídas (IID), com média 0 e variância σ_ε^2 .

2.3.2 Função Autocorrelação

A função ACF, do inglês *Autocorrelation Function*, é utilizada para identificar a correlação que uma determinada série temporal apresenta com ela própria, desfasada em k períodos e é utilizada como indicador estatístico das características dos dados.

A função ACF é definida por:

$$\rho_k = \frac{cov(Y_t, Y_{t-k})}{V(Y_t)}, k = 1, \dots, n - 1 \quad (2.9)$$

E pode ser estimada por:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, k = 1, \dots, n - 1$$

Onde Y_t corresponde ao valor observado no instante t , \bar{Y} ao valor médio dessas observações e n ao número de observações disponíveis.

A função de autocorrelação ACF, ρ_k dá-nos a correlação entre duas variáveis espaçadas por um tempo k (desfasamento/lag), ao caracterizar o desenvolvimento de Y_t ao longo do tempo, demonstrando o quão forte o valor observado, hoje, está correlacionado com os valores observados no passado e como os choques, hoje afetam valores futuros da variável estocástica. Além de nos ajudar a descrever os dados, a ACF também nos ajuda a verificar a estacionariedade, escolher modelos e fazer diagnósticos sobre regressões ou até a identificar a presença de sazonalidade. (Enders, 2009)

Da observação do correlograma amostral (gráfico $\hat{\rho}_k$ contra o lag k , ver exemplo figuras 4, 5 e 6) pode-se concluir, por exemplo, que a série apresenta correlação de curta duração se as primeiras correlações forem elevadas, mas descenderem rapidamente para zero, isto é, cada observação da série apenas se relaciona com as observações mais recentes – Figura 4. Numa série com dependência de longa duração, os valores nos *lags* só vão começar a decrescer ao fim de algumas observações – Figura 5. Por outro lado, se uma série apresentar sazonalidade, o correlograma irá apresentar valores elevados a cada intervalo de tempo correspondente ao período sazonal – Figura 6.

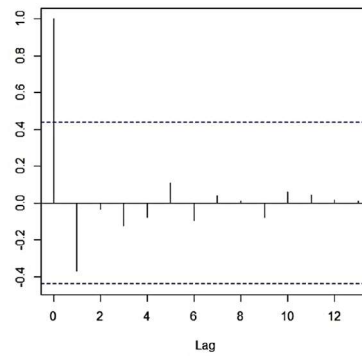


Figura 4 - Série com correlação de curta duração

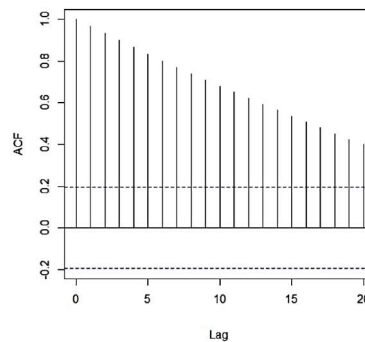


Figura 5 - Série com dependência de longa duração

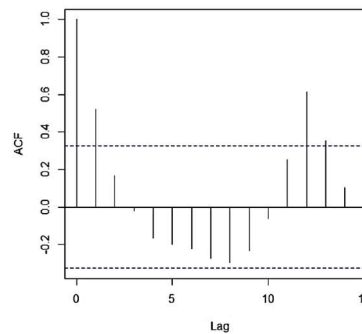


Figura 6 - Série com sazonalidade

2.3.3 Função Autocorrelação Parcial

As autocorrelações para intervalos sucessivos são intrinsecamente dependentes, isto é, se o primeiro elemento da série está fortemente relacionado com o segundo, e o segundo com o terceiro, então o primeiro está de alguma forma também relacionado com o terceiro. Para se obter uma informação sobre autocorrelações na série sem esta influência em cascata, é introduzido um outro conceito: função de autocorrelação parcial.

A autocorrelação parcial de atraso k corresponde à autocorrelação entre Y_t e Y_{t-k} que não é explicada pelos atrasos entre 1 a k :

$$\varphi_{kk} = \text{corr}(Y_t, Y_{t-k} | Y_{t-1}, \dots, Y_{t-k+1}) \quad (2.10)$$

A função de autocorrelação parcial, PACF, dá-nos a correlação entre a variável no instante t e um dos seus desfasamentos, retirando os efeitos de outros desfasamentos. Ou seja, define a correlação entre as observações Y_t e Y_{t-k} removendo o efeito das observações entre Y_{t-k} e Y_t . Esta remoção é feita determinando Y_k e Y_{t-k} como combinação linear das observações $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$ (Cryer e Chan, 2008).

3. Metodologia

No âmbito desta dissertação, propõe-se responder à questão:

- Qual o modelo de previsão a adotar pelo sistema de gestão de stocks para prever a procura de peças na MCozinha Peças?

Neste capítulo serão apresentadas as diferentes metodologias consideradas para a posterior descrição e análise dos dados estudados nesta dissertação e fornecidos no capítulo 5.

Assim será descrita a metodologia subjacente ao modelo ARIMA, o método de alisamento exponencial, o método de Holt-Winters aditivo e posteriormente o método de Croston. Nesta sequência definem-se os casos particulares do modelo ARIMA, a saber: AR, MA e ARMA.

Será utilizado o *software* R para fazer a estimação da previsão de cada um dos modelos.

3.1 Processos Autorregressivos

A série temporal Y_t é um processo autorregressivo de ordem p , denotado por $AR(p)$, se:

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t = \mu + \sum_{i=1}^p \varphi_i Y_{t-i} + \varepsilon_t \quad (3.11)$$

$$\varepsilon_t \sim IID(0, \sigma_\varepsilon^2) \quad (3.12)$$

Onde $\varphi_1, \dots, \varphi_p$ são constantes fixas ($\varphi_p \neq 0$), designadas por parâmetros de autoregressão, e $\{\varepsilon_t\}$ é o ruído branco: conceito já definido anteriormente.

Considerando o operador de atraso B , onde $BY_t = Y_{t-1}$, a representação toma a forma:

$$\triangleright \varphi(B)Y_t = Y_t - \varphi_1 BY_t - \varphi_2 B^2 Y_t - \dots - \varphi_p B^p Y_t = \mu + \varepsilon_t$$

$$\triangleright \varphi(B)Y_t = (1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)Y_t = \mu + \varepsilon_t$$

$$\triangleright \varphi(B)Y_t = \mu + \varepsilon_t$$

Por definição, a equação característica é dada pela equação $\varphi(B) = 0$ onde B é convencionalmente tratado como um número (real ou complexo).

Todas as raízes da equação característica terão de ter valor absoluto superior a 1, isto é, fora do círculo unitário para que o processo seja estacionário. (Cowpertwait e Metcalfe, 2009). Quando num AR(1), temos $\varphi = 1$, estamos na presença de uma raiz unitária, a qual basicamente, é sinónimo de não estacionariedade. O nome surge do facto da solução da equação característica correspondente ter uma raiz igual a 1. Testes de estacionariedade baseiam-se na ideia da verificação da existência de raiz unitária.

Características dos processos autorregressivos:

- Estacionário se as raízes de $\varphi(B) = 0$ estão fora do círculo unitário (soluções de $\varphi(z) = 0$)
- FAC: ρ_k tem um decrescimento exponencial ou sinusoidal amortecido para zero
- FACP: $\varphi_{kk} = 0$ se $k > p$ (corte para zero)

O modelo de previsão é dado pela seguinte expressão:

$$\hat{Y}_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} \quad (3.13)$$

Estes modelos também podem ser utilizados para modelar processos não estacionários.

3.2 Média Móvel

A série temporal Y_t diz-se ser um processo de médias móveis de ordem q , denotado por MA(q) se:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} = \mu + \sum_{i=0}^q \theta_i \varepsilon_{t-i} \quad (3.14)$$

$$\varepsilon_t \sim IID(0, \sigma_\varepsilon^2) \quad (3.15)$$

Em que Y_t é o processo, μ é a média do processo, ε_t é o ruído branco e q é a ordem da média móvel. Os parâmetros $\theta_1, \dots, \theta_p$ são constantes fixas ($\theta_q \neq 0$) e $\theta_0 = 1$. Tal como no processo autorregressivo, é possível escrever a equação anterior como um polinómio de ordem q em termos do operador de atraso B , onde:

$$B\varepsilon_t = \varepsilon_{t-1}, B^2\varepsilon_t = \varepsilon_{t-2}, \dots, B^q\varepsilon_t = \varepsilon_{t-q} \quad (3.16)$$

Obtendo-se:

$$Y_t = \mu + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t = \mu + \theta(B) \varepsilon_t \quad (3.17)$$

Dado que os processos de médias móveis são uma soma finita de termos de um ruído branco então são estacionários e, por consequência, possuem média e auto-covariância independentes do tempo. Nesta sequência concluiu-se que a média do processo MA(q) é μ , visto que os termos são mutuamente independentes, a variância é dada por: $\sigma_\varepsilon^2(1 + \theta_1^2 + \dots + \theta_q^2)$ (Cowpertwait e Metcalfe, 2009).

Características:

- Sempre estacionário
- FAC: $\rho_k = 0$ se $k > q$ (corte para zero, tende para zero)
- FACP: φ_{kk} tem um decrescimento exponencial ou sinusoidal amortecido para zero

O modelo de previsão do MA é dado pela seguinte expressão:

$$\hat{Y}_t = \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \mu \quad (3.18)$$

3.3 Processos Autorregressivos e de Médias Móveis

Uma série temporal Y_t constitui um processo autorregressivo de ordem p e de médias móveis de ordem q , denotado por ARMA (p, q) , se:

$$Y_t = \mu + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3.19)$$

$$Y_t = \sum_{i=1}^p \varphi_i Y_{t-i} + \mu_t + \sum_{i=0}^q \theta_i \varepsilon_{t-i} \quad (3.20)$$

$$\varphi(B)Y_t = \mu + \theta(B)\varepsilon_t \quad (3.21)$$

Os processos AR e MA são assim combinados: onde p é o grau de AR, e q representa o grau de MA. Se $p = 0$ então o processo é MA(q); se $q = 0$ então o processo é AR(p). Os coeficientes p e q são escolhidos através da análise da autocorrelação global e parcial.

Assim, é inicialmente medida para cada desfasamento de k , a relação entre Y_t e Y_{t-k} , considerando também as variáveis $Y_{t-1}, \dots, Y_{t-k+1}$. E em seguida, a relação entre Y_t e Y_{t-k} , sem considerar as outras variáveis. A autocorrelação global e parcial são analisadas pelos respectivos correlogramas e os graus p e q que têm de ser utilizados estão ligados à distribuição mostrada pelos correlogramas.

O modelo de previsão do ARMA é dado pela seguinte expressão:

$$\hat{Y}_t = \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \mu \quad (3.22)$$

Note-se as seguintes características:

- O processo é estacionário se todas as raízes de $\varphi(B) = 0$ estão fora do círculo unitário;
- FAC: ρ_k tem um decrescimento exponencial ou sinusoidal amortecido para zero
- FACP: φ_{kk} tem um decrescimento exponencial ou sinusoidal amortecido para zero

O método ARMA é mais adequado quando as séries temporais são estacionárias.

3.4 Processo Autorregressivo Integrado de Média Móvel

Um processo autorregressivo integrado de média móvel é uma generalização dos processos ARMA que incorpora séries não estacionárias. Assim, tal como nos métodos anteriores, possui apenas uma variável e descreve o comportamento de uma variável em termos dos seus valores passados. (Brockwell e Richard A. Davis, 2002).

Os modelos ARIMA, elaborados por Box e Jenkins em 1970, normalmente são representados por ARIMA (p, d, q) . O parâmetro p refere-se ao número de termos Autorregressivos; o parâmetro d diz respeito ao número de diferenciações que são necessárias para transformar a série não estacionária, em estacionária, conceito que será abordado de seguida e por fim o q trata-se do número de médias móveis. O modelo ARIMA resulta assim na junção de três componentes: a componente Autorregressiva (AR), a componente da diferenciação (I) e ainda a componente das Médias Móveis (MA).

Como se referiu anteriormente, a estacionariedade de uma série temporal implica que esta tenha média e variância constantes e covariância invariante do tempo, dependendo apenas do desfasamento temporal. Assim, a diferenciação é utilizada para diferenciar a série, isto é, desfasar a série temporal em determinado número de instantes tendo como objetivo estabilizar a média da série. Então, dada uma série temporal Y_t , a primeira diferença desta é dada pela diferença entre valores consecutivos: Y_t e Y_{t-1} .

$$\Delta Y_t = Y_t - Y_{t-1} \quad (3.23)$$

Para facilitar a visualização das expressões matemáticas é usado o operador de atraso, B , o qual atrasa as séries temporais em d instantes, ou seja: $B^d Y_t = Y_{t-d}$, resultando a primeira diferença em $\Delta Y_t = (1 - B)Y_t$.

Se ainda assim, após a primeira diferenciação, a média da série temporal não estiver estabilizada, volta-se a diferenciar até que esta fique estacionária. A diferenciação de 2ª ordem é dada por:

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2} = (1 - B)^2 Y_t \quad (3.24)$$

Depois da série estacionária ter sido diferenciada d vezes, $(1 - B)^d Y_t$, sendo $d \geq 1$, poder-se-ia aplicar um modelo ARMA (p, q) . Consequentemente, a expressão do modelo ARIMA (p, d, q) resulta em:

$$(1 - \varphi_1 B - \dots - \varphi_p B^p)(1 - B)^d Y_t = \mu + (1 - \theta_1 B - \dots - \theta_q B^q)\varepsilon_t \quad (3.25)$$

$$\varphi(B)(1 - B)^d Y_t = \mu + \theta(B)\varepsilon_t \quad (3.26)$$

Face ao referido, um processo ARIMA é uma generalização do modelo ARMA, adicionando a diferenciação de ordem $d \geq 0$, para uma possível estabilização da série. Este modelo é aplicado em casos em que não há uma evidência da estacionariedade.

Modelo previsão do ARIMA é dado pela seguinte equação:

$$\widehat{\Delta^d Y_t} = \varphi_1 \Delta^d Y_{t-1} + \dots + \varphi_p \Delta^d Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \mu \quad (3.27)$$

Numa situação em que um ou dois parâmetros sejam nulos, o modelo a utilizar baseia-se no parâmetro diferente de zero, como por exemplo um modelo Autorregressivo, ARIMA $(1,0,0)$, que equivale a um modelo Autorregressivo de 1ª ordem ou $AR(1)$.

$$ARIMA(p, 0, 0) = AR(p)$$

$$\text{ARIMA}(0,0,q) = \text{MA}(q);$$

$$\text{ARIMA}(p,0,q) = \text{ARMA}(p,q)$$

Em síntese, para aplicar um modelo ARIMA, será necessário:

1. Transformar a série temporal numa série estacionária, sendo necessário diferenciar a série d vezes; este processo pode ser realizado visualmente ou aplicando algum teste da raiz unitária que permita estimar o valor de d ;
2. Estimar os parâmetros p e q , através da análise do gráfico da ACF.

Como nem sempre é fácil estimar estes valores dada a complexidade dos cálculos é vantajoso recorrer a algumas ferramentas disponíveis, como o *R* (R Core Team, 2014), pois conseguem de forma automática estimar e devolver os parâmetros ótimos a considerar. (Hyndman and Khandalar, 2008)

Os dois principais métodos de estimação dos parâmetros do ARIMA são o (i) método da máxima verosimilhança⁵ e o (ii) método dos mínimos quadrados⁶. (Box *et al*, 1994)

Nesta modelação estão incluídas cinco etapas: identificação de modelos, estimação dos parâmetros dos modelos, avaliação do diagnóstico, seleção do melhor modelo e, por último, se o modelo for adequado, a previsão. (Caiado, 2012)

3.4.1 SARIMA

Os modelos ARIMA são, por norma, representados por ARIMA (p, d, q) . No entanto, existem também os modelos sazonais que são escritos como ARIMA $(p, d, q) (P, D, Q)m$. Assim, p, d, q têm o mesmo significado de P, D e Q , só que as letras minúsculas correspondem à parte não sazonal e as letras maiúsculas à parte sazonal.

A diferença será na etapa da identificação que consiste na estimação dos valores para p, d, q , e P, D e Q , no caso de haver sazonalidade, para que se escolha o modelo ARIMA $(p, d, q) (P, D, Q)m$ que melhor se ajusta à série temporal. Após serem identificados os

⁵ Permite determinar os valores dos parâmetros do modelo que maximizam a probabilidade de ocorrência de um conjunto de observações semelhantes àquelas de que efetivamente se dispõe. Para obter as estimativas desses parâmetros é utilizado um processo iterativo no qual se maximiza a função de verosimilhança associado ao modelo.

⁶ Método que procura encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados.

modelos candidatos, efetua-se a estimação dos parâmetros tal como foi demonstrado anteriormente no modelo ARIMA.

3.5 Métodos de alisamento exponencial

Os métodos de alisamento exponencial foram desenvolvidos por Brown (1963), Holt (1957) e Winters (1960) são métodos de previsão que utilizam combinações ponderadas das observações passadas para posteriormente preverem valores futuros.

As previsões destes métodos de alisamento exponencial são produzidas a partir de fórmulas simples tendo em conta a tendência e efeitos sazonais presentes nos dados a analisar. Este tipo de procedimentos é amplamente usado como técnicas de previsão em gestão de inventários e para estimar consumos. O termo alisamento exponencial indica que os pesos das observações são exponencialmente decrescentes com a antiguidade das mesmas, ou seja, quanto mais recente é a observação, maior é o seu respetivo peso (Gardner, 2006; Caiado, 2011; Hyndman, 2012).

Hyndman (2012) refere ainda que tanto o ARIMA como o Alisamento Exponencial, só usam a própria série temporal para realizar as previsões, ou seja, constituem modelos univariados.

3.5.1 Métodos de alisamento exponencial simples

Um dos métodos mais simples, o Alisamento Exponencial Simples, do inglês *Exponential Smoothing* (ETS), é definido como uma combinação linear de todas as observações conhecidas, dando mais peso às observações mais recentes. Este método é usado para séries temporais sem tendência.

Segundo Makridakis *et al.* (1998), o alisamento exponencial simples considera a previsão correspondente ao período anterior e faz um acerto para obter a previsão para o período seguinte. O acerto é realizado através da multiplicação do erro de previsão do período anterior por uma constante de alisamento $0 < \alpha < 1$, e quanto maior for α maior será a sua resposta ao impulso. Isto é, se para α for escolhido um valor próximo de 1, as últimas observações irão ter maior peso e as estimativas irão adaptar-se mais rapidamente a variações nas características da série. Já os valores mais próximos de 0 dão origem a sequências de estimativas mais estáveis, visto a nova previsão ser muito próxima da previsão anterior.

A representação matemática do modelo é expressa pela equação:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t) \quad (3.28)$$

Onde \hat{Y}_{t+1} é a previsão para o período $t + 1$, \hat{Y}_t é a previsão para o período t com origem no período $t - 1$, Y_t a procura no período anterior e α a constante de alisamento.

3.5.2 Método de Croston

Croston (1972) notou que a aplicação do alisamento exponencial simples (ETS) a séries que descrevem padrões de procura com períodos nulos dava origem a níveis de stock inadequados, pois observa-se a colocação de um maior peso nas observações mais recentes, levando a que a estimativa da previsão seja mais alta, logo após haver procura e mais baixa antes desta ocorrer. Sugere assim um método no qual se aplica um ETS sobre os elementos não nulos da série temporal e separadamente outro ETS sobre os intervalos entre os elementos diferentes de zero da série temporal.

Assim, Croston prevê separadamente o tempo entre observações consecutivas P_t e a magnitude individual destas observações Z_t . Consequentemente isto traduz-se em que a aplicação do ETS, para séries irregularmente espaçadas, seja feita quer à série dos intervalos de tempo $\{\Delta P_t\}_{t=1,\dots,N}$ quer à série das quantidades $\{Z_k\}_{t=1,\dots,N}$, de modo a serem obtidas as previsões um passo à frente.

Os parâmetros de alisamento das duas aplicações do ETS são considerados iguais e são denotados por α . No período t em análise, se a procura não ocorrer, a magnitude individual dos estimadores P_t e Z_t mantém-se inalterada. Como $Y_t = 0$, logo:

$$P_t = P_{t-1} \quad (3.29)$$

$$Z_t = Z_{t-1} \quad (3.30)$$

$$q = q + 1 \quad (3.31)$$

Se por sua vez a procura ocorrer, $Y_t > 0$, as estimativas são atualizadas por:

$$P_t = \alpha(q - P_{t-1}) + P_{t-1} \quad (3.32)$$

$$Z_t = \alpha(Y_t - Z_{t-1}) + Z_{t-1} \quad (3.33)$$

$$q = 1 \quad (3.34)$$

Onde,

Y_t Procura/valor observado no período t

P_t Estimativa alisada do nº médio de períodos de tempo entre os elementos não nulos da série

Z_t Estimativa alisada do tamanho médio de uma procura não nula

q Número de períodos de tempo desde a última observação diferente de zero

α Constante de alisamento que assume valores compreendidos entre 0 e 1

Combinando as estimativas do tempo entre observações não nulas e a magnitude individual dessas observações, a previsão da procura no período t é dada como:

$$\hat{Y}_t = \frac{Z_t}{P_t} \quad (3.35)$$

Estas estimativas apenas são atualizadas quando a procura ocorre. Quando existe procura em cada período, o método Croston é idêntico ao método ETS, ou seja:

$$\hat{Y}_t = Z_t \quad (3.36)$$

3.5.3 Método de Holt

Para as séries que apresentarem tendência, Holt (1957), desenvolveu um modelo alargado do alisamento exponencial simples. Neste modelo é incorporada a componente tendência para a sua previsão, como se pode verificar na seguinte equação:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (3.37)$$

Onde Y_t é o valor observado no período t , L_t é a estimativa para o nível da série no período t e T_t é a estimativa do declive (tendência) da série no mesmo período t . A equação ajusta diretamente L_t através da tendência do período anterior, T_{t-1} , adicionando o último valor alisado L_t .

A componente tendência, T_t , é obtida a partir da seguinte equação:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (3.38)$$

O método desenvolvido recorre ao uso de duas constantes do alisamento, α e β (ambas com valores entre 0 e 1). Os valores de α e β são encontrados por tentativas sobre a série histórica, sendo frequente seleccionar aqueles que minimizam o erro quadrático médio.

O nível é inicializado, tomando do $L_1 = Y_1$ e a tendência é inicializada por $T_1 = Y_2 - Y_1$

Dado que as estimativas das componentes da série são funções de valores passados, exigem que, para o arranque do cálculo das previsões, sejam definidos valores iniciais para cada uma das componentes. Leva a que a importância dos valores iniciais das componentes resulte da natureza recursiva das equações de atualização.

A previsão para \hat{Y}_{t+m} , é obtida através da soma do nível amortecido com a componente tendência, em que m representa o número de períodos para qual é escolhida a previsão:

$$\hat{Y}_{t+m} = L_t + m.T_t \quad (3.39)$$

3.5.4 Método de Holt-Winters

O método de Holt foi, em 1960, difundido por Winters para aplicar em séries temporais que além de apresentarem tendência, apresentam igualmente sazonalidade. Apesar do método de Holt-Winters ser bastante idêntico ao método de Holt, este adiciona ainda uma equação que se relaciona com a sazonalidade. É também um método de alisamento exponencial e permite construir previsões, considerando o crescimento, a tendência e a sazonalidade. (Makridakis *et al.*, 1998). Desta forma, o método Holt-Winters, estima três parâmetros de alisamento para cada uma das componentes apresentadas: nível, tendência e sazonalidade. A variação sazonal pode ser aditiva (modelo aditivo) ou multiplicativa (modelo multiplicativo) (Tratar, 2010).

Assim, o método de Holt-Winters apresenta dois modelos distintos de cálculo, o primeiro que modela a sazonalidade de maneira aditiva e o segundo modelo, que modela a sazonalidade de forma multiplicativa.

A estimação de cada uma das componentes é efetuada de acordo com as seguintes equações recursivas, já vistas anteriormente com o acréscimo da equação da sazonalidade:

$$\text{Nível: } L_t = \alpha(Y_t - S_{t-L}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (3.40)$$

$$\text{Tendência: } T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (3.41)$$

$$\text{Sazonalidade: } S_t = \gamma(Y_t - T_t) + (1 - \gamma)S_{t-L} \quad (3.42)$$

Onde, para além da notação definida anteriormente:

S_t	Estimativa do fator sazonal no período t
α	Constante de amortecimento do nível, com $0 \leq \alpha \leq 1$
β	Constante de amortecimento da tendência, com $0 \leq \beta \leq 1$
γ	Constante de amortecimento do fator sazonal, com $0 \leq \gamma \leq 1$
L	Número de unidades de tempo incluídas em cada ciclo de sazonalidade.

A função de previsão do modelo aditivo de Holt-Winters (para séries temporais com duração de período p) é dada por:

$$\hat{Y}_{t+k} = L_t + kT_t + S_{t+k-mL} \quad (3.43)$$

$\hat{Y}_t(k)$	Previsão efetuada no período t para o valor da série no período $t + k$
m	Número de ciclos de sazonalidade completos entre os períodos t e $t + k$.

Tendo em conta, as equações recursivas, as estimativas das componentes da série são função das observações passadas, das constantes de amortecimento e dos valores iniciais das componentes. Tal como no método de alisamento exponencial simples, constantes de amortecimento destes modelos são parâmetros que permitem atribuir uma maior ou menor ponderação aos valores presentes ou passados: a constantes de amortecimento mais elevadas corresponde uma maior ponderação dos valores mais recentes, tornando o modelo mais reativo embora com maior instabilidade.

Esta formulação recursiva apresentada reporta-se ao trabalho original de Winters (1960). No entanto, é frequente representar-se este modelo numa forma equivalente, onde

se explicita a alteração de cada componente com base numa proporção do último erro cometido, $e_t = Y_t - \hat{Y}_t$.

$$L_t = L_{t-1} + T_{t-1} + \alpha \frac{e_t}{S_{t-L}} \quad (3.44)$$

$$T_t = T_{t-1} + \alpha\beta \frac{e_t}{S_{t-L}} \quad (3.45)$$

$$S_t = S_{t-L} + \gamma(1 - \alpha) \frac{e_t}{L_t} \quad (3.46)$$

Esta especificação do modelo evidencia que cada estimativa é obtida corrigindo a estimativa anterior, sendo a correlação tanto maior quanto maior for o erro e as constantes de amortecimento.

3.6 Erros de previsão

Em cada um dos métodos descritos anteriormente, a previsão da procura de um produto é gerada pela interação de um número de fatores bastante complexos para serem descritos com precisão. Desta forma, todas as previsões contêm uma parcela de erro ou resíduo no resultado apresentado. E esses erros, segundo os autores Chase *et al.* (2006) podem ser de dois tipos: erros sistemáticos ou erros aleatórios. Os erros sistemáticos correspondem aos erros cometidos de forma consistente, como por exemplo: o uso de relações erradas entre variáveis ou utilização errada da linha de tendência ou uma mudança errada na procura sazonal. Os erros aleatórios são os erros que não podem ser explicados pelo modelo de previsão usado, sendo o melhor modelo aquele que minimiza as medidas de erro. Com o intuito de comparar os diferentes métodos e modelos de previsão serão utilizadas três medidas de erro: MSE, MSE e MAPE.

Passamos a considerar as seguintes variáveis, onde H representa o número de observações disponíveis para teste, Y_t o valor observado no instante t e \hat{Y}_t a previsão para o instante t . (Caiado, 2011; Ramos, 2012)

MSE, *Mean Square Error* ou erro quadrático médio, assinala um desvio sistemático por defeito ou por excesso e é dado pela equação:

$$MSE = \frac{1}{H} * \sum_{t=1}^H (Y_t - \hat{Y}_t)^2 \quad (3.47)$$

De referir que o MSE é obtido através de uma escala diferente do original, neste caso o quadrado, sendo a sua principal desvantagem a sensibilidade a erros elevados.

Por outro lado, o *Mean Absolute Error* ou erro absoluto médio é obtido da através da equação:

$$MAE = \frac{1}{H} * \sum_{t=1}^H |Y_t - \hat{Y}_t| \quad (3.48)$$

Tanto o MSE, como o MAE são úteis sobretudo para comparar métodos aplicados a uma mesma série temporal, no entanto o MAE não amplifica os erros mais elevados ao contrário do MSE.

MAPE, *Mean Absolute Percentage error* ou erro percentual absoluto médio, da previsão (*forecast*) em percentagem, isto é, mede o tamanho do erro em termos percentuais. É uma medida relativa que não depende da escala de dados, já que divide a diferença ($Y_t - \hat{Y}_t$) pelo valor atual (Y_t). (Davydenko e Fildes, 2013)

$$MAPE = \frac{1}{H} * \sum_{t=1}^H \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| * 100 \quad (3.49)$$

3.7 Seleção do modelo

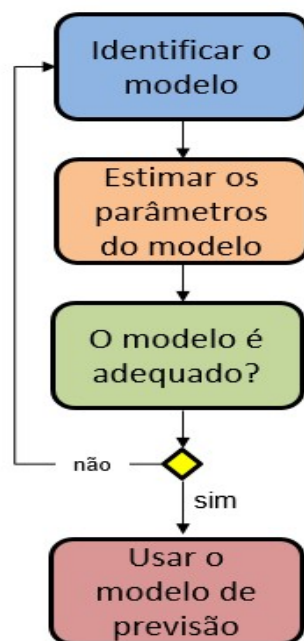
Um dos problemas da previsão dos modelos de previsão consiste em determinar qual o método mais adequado para cada caso. O método de Box-Jenkins utiliza um algoritmo matemático complexo onde vários modelos são avançados e em que a partir das observações históricas, se calculam diferentes estatísticas, para com base nelas se identificar o modelo mais adequado para cada série temporal.

Segundo Box *et al.* (1994), são três as fases de construção de um modelo Box-Jenkins:

- i. identificação do modelo que melhor se adapta a cada série;
- ii. estimação dos parâmetros do modelo identificado na primeira fase;
- iii. diagnóstico: verificação, no qual se avalia se o modelo proposto é adequado, através por exemplo, da diferença do valor atual da série e o valor apresentado pelo modelo em análise (resíduos).
- iv. previsão

A estimação dos parâmetros do modelo identificado faz-se visualmente, recorrendo ao método dos mínimos quadrados ou da máxima verosimilhança. De seguida, depois de se estimarem os parâmetros do modelo é necessário verificar se existe um desajuste pronunciado entre aquele modelo e a série real, para isso são utilizados vários tipos de testes.

Os parâmetros ótimos são devolvidos através de critérios de informação que ajudam a medir a qualidade do ajustamento entre os vários modelos possíveis para determinada série temporal. O melhor modelo será aquele que otimizar o valor de cada um destes critérios. Estes critérios comparam os vários modelos estatísticos construídos com base nos dados e utilizando diferentes números de parâmetros no modelo. A figura 7 resume esquematicamente os passos do método.



Box *et al.* (1994)

Figura 7 – Diagrama Box-Jenkins

4. Caso de estudo

O negócio automóvel, nas suas diversas áreas, cresceu substancialmente, criou uma cultura própria e desenvolveu necessidades às quais o Grupo MCoutinho, com uma postura inflexível de integridade, rigor e inovação, tem respondido, com a apresentação de novas marcas e a criação de empresas e serviços que refletem bem a dedicação partilhada pelo fundador.

O Grupo MCoutinho com 16 Marcas de distribuição exclusiva, representando mais de 65% do mercado, com 3,2% de Quota de mercado nacional, com uma implantação geográfica alargada com 46 pontos de venda e 64 reparadores autorizados, pretende ser reconhecido como uma referência nos mercados onde atua, aspirando a ser uma das opções preferenciais a considerar pelos clientes do sector automóvel.

A MCoutinho Peças tem a missão de ser a empresa que melhor entende e satisfaz as necessidades de produtos no sector de peças automóveis e rege-se pelos valores de profissionalismo, transparência e dedicação.

Em 2009, foi inaugurado, no distrito de Lisboa, o segundo Centro Logístico, reforçando as operações do Grupo, até então concentradas no Porto, e o posicionamento como operador com cobertura nacional, proporcionando um serviço logístico de excelência e dando resposta às exigências do mercado.

Em 2016 a MCoutinho Peças lançou um novo Portal que permite pesquisar e encomendar peças de forma rápida e intuitiva, além de disponibilizar uma série de ferramentas únicas no mercado em modo Self Service, da gestão de stocks à possibilidade de elaborar orçamentos. O novo Portal permite ainda verificar o estado das encomendas e respetivas previsões de entrega, extrato de conta e duplicados de faturas até estatísticas de evolução de compras.

Atualmente, a MCoutinho Peças é líder de mercado em Peças Originais em Portugal, representando 27 marcas automóveis e atingindo um volume de negócios anual superior a 34 milhões de euros. Proporciona, também, uma gama multimarca verdadeiramente abrangente - mais de 95% do parque automóvel nacional - num modelo “uma chamada, uma fatura, uma entrega”.

A oferta da MCoutinho Peças consubstancia-se, nas seguintes valências: num parceiro de referência, líder de mercado no abastecimento de peças originais; numa gama completa multimarca; numa proximidade e flexibilidade no desenvolvimento do negócio; num serviço logístico flexível, reconhecido como sendo a referência no mercado; numa forte aposta em sistemas e tecnologia de informação materializando-se na gestão de stocks e sua disponibilidade.

Acresce também referir que atualmente o Grupo MCoutinho está a desenvolver um conjunto de projetos que o posicionarão dentro das melhores práticas empresariais ambientalmente sustentáveis, com a criação do conceito "Planeta MCoutinho", cuja missão é apoiar a proteção ambiental e o desenvolvimento social, esforçando-se por utilizar os recursos naturais de forma eficiente, favorecer a utilização de fontes energéticas renováveis e promover o recurso a materiais recicláveis. Este desafio implica o compromisso de toda a equipa MCoutinho, que atualmente integra mais de 600 colaboradores.

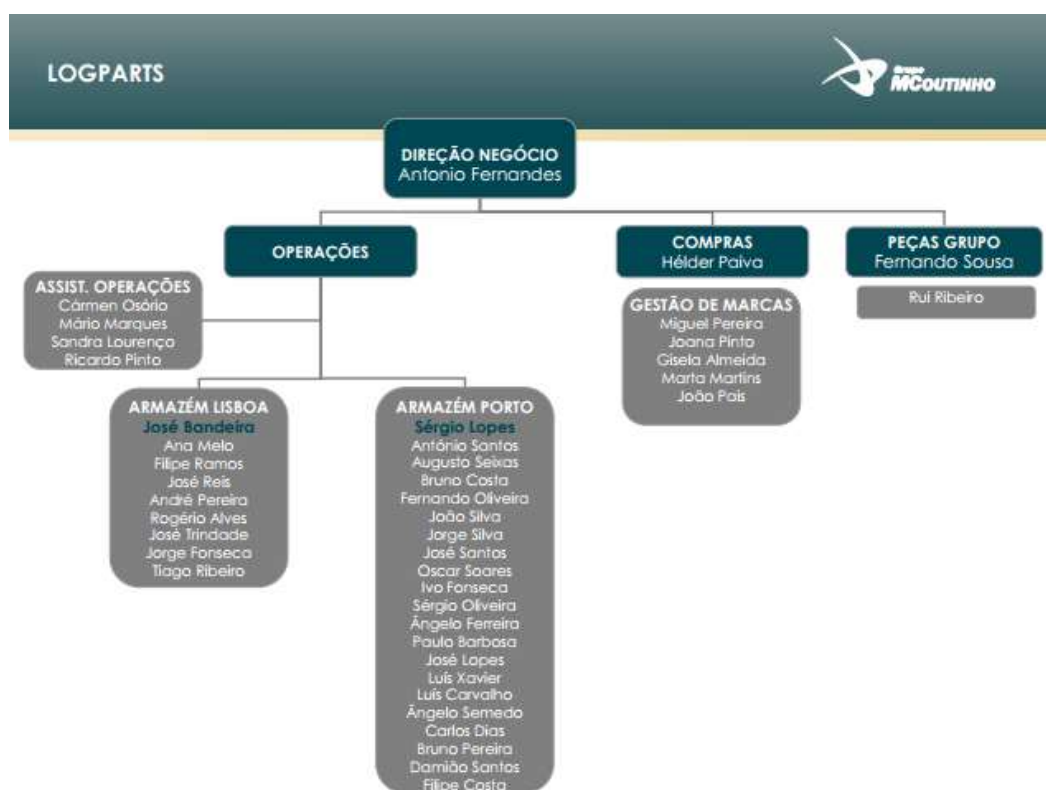


Figura 8 – Organograma Grupo MCoutinho

4.1 Descrição da amostra

Para a realização do projeto, o Grupo MCoutinho disponibilizou no âmbito deste trabalho o histórico das vendas diárias de janeiro de 2010 até agosto de 2016.

Dentro das 27 marcas de automóveis representadas pela MCoutinho, a amostra irá recair sobre a marca Peugeot que detêm um volume de faturação de sensivelmente 4 milhões de euros anuais, detendo aproximadamente 8000 artigos em stock.

Analizamos inicialmente e descritivamente a série temporal de uma referência tipo, passando posteriormente à análise global de todas as referências.

4.2. Modelação e análise quantitativa

O caso de estudo desta dissertação pretendeu efetuar a previsão do próximo consumo com base nos dados dos consumos do armazém do Porto. O objetivo é conseguir melhorar o sistema de previsões de vendas, através de um sistema avançado de previsões.

Em termos computacionais, a modelação e as previsões foram efetuadas utilizando o *software R*⁷, em conjunto com o ambiente de programação *RStudio*⁸, cujo *layout* se observa na figura 8, que permite facilmente a instalação de *packages*⁹ criados pela comunidade. O *package* a ser utilizado, maioritariamente, nesta dissertação para a modelação e previsão denomina-se “*forecast*”. Este *package* permite fazer a modelação dos dados utilizando modelos ARIMA, ETS, Holt-Winters e Croston, assim como a sua previsão, elaborando ainda os cálculos dos erros de previsão. Da mesma forma, este foi o *software* utilizado na preparação e tratamento dos dados.

No essencial, pretende-se perceber se os modelos são adequados e quais os melhores para cada contexto de artigos.

⁷ R é uma linguagem de programação e ambiente para computação estatística. Software em evolução constante, grátis e com uma ampla contribuição de vários programadores/comunidade. R fornece uma ampla variedade de técnicas estatísticas e gráficas (modelação não-linear, testes estatísticos clássicos, análise de séries temporais linear, ...) e é altamente extensível. (R Core Team, 2014)

⁸ RStudio é uma aplicação para o *software R*, um ambiente de desenvolvimento integrado grátis e de código aberto, que possui uma interface gráfica mais amigável do que o ambiente de programação disponibilizado pelo R.

⁹ Package: consiste num conjunto de funções que quando invocadas executam rotinas de programação. (R Core Team, 2014)

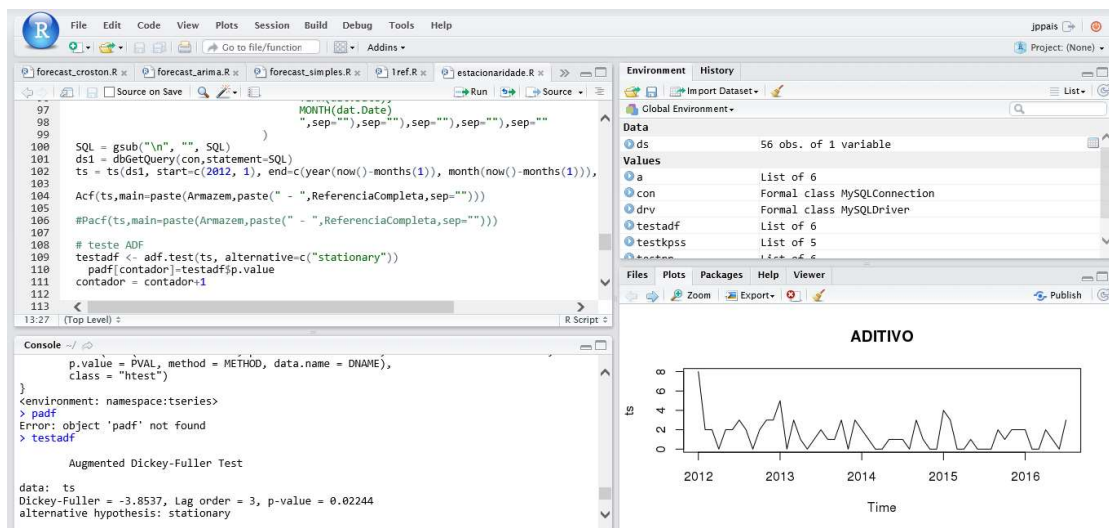


Figura 9 – Ambiente de modelação e previsão em RStudio

4.3 Recolha e preparação dos dados

Devido ao grande volume de dados diários e ao uso recorrente de dados mensais em trabalhos presentes na literatura, para uma análise mais fácil decidiu-se fazer uma agregação mensal das vendas com sensivelmente 80 observações por série de dados. Assim, para prever agosto de 2016, dispomos de uma matriz com 80 linhas relativas aos dados desde janeiro de 2010 até julho de 2016.

Na análise descritiva, serão retiradas as últimas observações e serão ajustados os seguintes modelos para cada série:

- Modelo Holt-Winters sem sazonalidade e sem tendência (alisamento exponencial simples);
- Modelo Holt-Winters com sazonalidade aditiva;
- Modelo Croston;
- Modelo ARIMA com (p, q, d) determinado através dos testes ADF, ACF e PACF.

Para cada modelo ajustado, será realizada a previsão de 1 passo à frente e multi-passo. Assim, pode-se comparar com as 3 observações originais das séries e medir qual o modelo com maior poder preditivo em termos de MSE, MSE e MAPE.

Depois de identificados os artigos destinados para o estudo deste projeto, foi feita uma recolha de dados, do consumo de cada material durante os anos de 2010 a 2016. A informação foi extraída através de uma ferramenta de trabalho amplamente utilizada na empresa, o *Quiter*. Foi tirada uma listagem que possui um arquivo de todos os movimentos do material e a respetiva data. Esta informação foi compilada e tratada numa folha de cálculo Excel.

4.4 Tratamento dos dados

4.4.1 Classificação dos artigos por classe

É necessário agora estudar e tratar os dados disponíveis de forma a adequá-los a cada série face as suas características. Como visto anteriormente, a análise ABC/Pareto; verifica-se, à medida que nos deslocamos da classe A para a B e depois para a C, o número de artigos vai aumentando e o valor financeiro vai diminuindo:

Nº artigos: Classe A < Classe B < Classe C

Valor financeiro: Classe A > Classe B > Classe C

Finalmente, a análise ABC permite detetar os artigos sem movimento, isto é, os artigos obsoletos, os quais devem ser retirados do stock global porque o seu armazenamento implica custos sem contrapartidas para o funcionamento da empresa. De notar que esta análise não é exclusiva dos problemas dos stocks, sendo utilizada em várias atividades da gestão empresarial.

No que diz respeito à análise da curva ABC, esta serve para analisar a dependência ou risco face a um determinado grupo de produtos, ou ainda para que tipo de produtos a organização se deve focar em função do valor financeiro investido. Consiste em ordenar os clientes por ordem decrescente de valor, de modo a poder segmentar por grau de dependência, de risco ou ainda por outro critério a definir. Os artigos da classe A devem merecer maior atenção e vigilância sendo necessário controlar a movimentação dos artigos deste grupo em referência.

A classe A do caso em estudo é constituída por 856 artigos, que significa aproximadamente 10% da totalidade dos artigos existentes em *stock* da marca. Aquele número de artigos corresponde a 914.000€, ou seja, a 80% do valor financeiro existente no *stock* global em análise, ver figura 10.

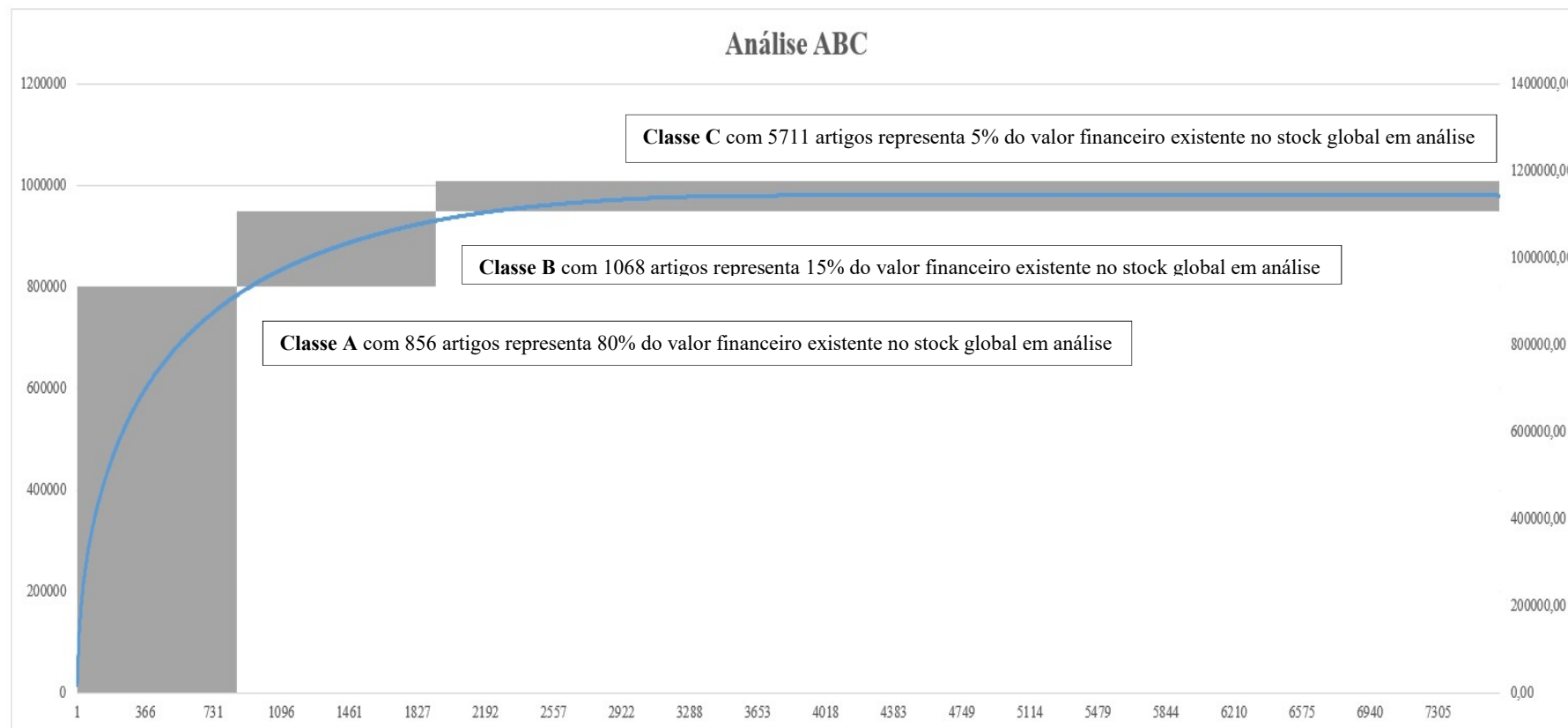


Figura 10 – Análise ABC de 7635 referências com consumo no ano de 2016.

4.4.2 Avaliação da estacionariedade

A escolha do modelo deverá ser feita tendo em conta um teste de estacionariedade e a análise das funções ACF e PACF. Uma série não estacionária deve ser diferenciada até que tal se torne estacionária. Para se perceber se a série observada é ou não estacionária, são apresentados nesta secção os testes mais comuns para testar a estacionariedade de um conjunto de observações. Nomeadamente:

1. Teste Dickey-Fuller Aumentado (ADF);
2. Teste Phillips-Perron (PP);
3. Teste Kwiatkowski-Phillips-Schmidt-Shin (KPSS).

A ideia subjacente a estes testes é a verificação da existência de raiz unitária¹⁰. (Hyndman, 2012)

Assim considere-se a série temporal Y_t escrita como:

$$Y_t = \varphi Y_{t-1} + \varepsilon_t \quad (5.50)$$

correspondente ao modelo $AR(1)$ e sujeito a $\varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$. Neste caso, se $|\varphi| < 1$ então Y_t é estacionário. A hipótese a testar é então $H_0: \varphi = 1$, ou seja, a série não é estacionária contra a hipótese alternativa $H_1: |\varphi| < 1$ (a série é estacionária) ou, conforme é mais comum, reescrevendo o modelo como:

$$\Delta Y_t = (\varphi - 1)Y_{t-1} + \varepsilon_t = \pi Y_{t-1} + \varepsilon_t \quad (5.51)$$

Testar-se $H_0: \pi = 0$ contra $H_1: \pi < 0$

Esta é a abordagem do teste Dickey-Fuller (DF) onde os erros ε_t são assumidos como sendo não correlacionados, ou seja, se esta hipótese não se verificar podemos aumentar o teste DF de forma a incluir um número maior de defasagens das diferenças. Isto é, teste de Dickey-Fuller Aumentado, conhecido na literatura como teste ADF (Augmented Dickey-Fuller) faz o estudo da estacionariedade sobre a seguinte regressão:

$$\Delta Y_t = \beta_0 + \beta_1 t + \pi Y_{t-1} + \sum_{i=1}^{p-1} \alpha_i \Delta Y_{t-i} + \varepsilon_t \quad (5.52)$$

¹⁰ Se a série tem raiz unitária então não é estacionária.

β_0 é a constante de interceção, também denominada como drift da série; β_1 é o coeficiente de tendência; π é o coeficiente de presença de raiz unitária e p é o número de lags tomados na série (Gujarati, 2004, secção 21.9).

A estatística de teste é:

$$T = \frac{\hat{\pi}}{sd(\hat{\pi})} \quad (5.53)$$

$\hat{\pi}$ é um estimador para π e $sd(\hat{\pi})$ é um estimador para o desvio padrão de π . Os valores críticos da estatística T foram tabelados por Dickley e Fuller (Gujarati, 2004) através de simulação Monte Carlo e variam nos casos em que há presença de uma constante de interceção, constante de tendência ou na presença de ambas.

Valores elevados do *p-value* sugerem não estacionariedade e valores pequenos sugerem estacionariedade.

O teste PP é uma generalização do teste DF admitindo o caso em que os erros são correlacionados e/ou heteroscedásticos e têm idênticas hipóteses H_0 e H_1 . Já o teste KPSS tem como hipóteses H_0 : a série é estacionária e H_1 : a série apresenta raiz unitária. Neste teste, as hipóteses diferem dos dois anteriores.

4.4.3 Implementação da gestão de stocks

Para realizar a otimização do processo de gestão de inventário, foram estudados os sistemas de controlo de *stocks*, referidos no capítulo 2. Conforme as necessidades da empresa verificou-se que um dos sistemas mais indicados a ser utilizado seria o método de revisão periódica, devido ao dia da encomenda de stock da Peugeot ser uma vez por semana numa data fixa. No entanto, existem marcas em que esta questão não se coloca, encomendas diárias de stock, e seria usado o método de revisão contínua. No entanto, nesses últimos casos a previsão não assume um fator tão importante e decisivo como no caso da Peugeot.

O modelo de revisão contínua é adequado quando se verifica pelo menos uma das seguintes condições: variação muito grande do consumo, preço de custo elevado, importância elevada e risco de rutura baixo e disponibilidade permanente de stock no fornecedor. Se a procura fosse perfeitamente estável, ou seja, se as entregas dos

fornecedores nunca contivessem peças defeituosas e os prazos de aprovisionamento fossem sempre precisos e cumpridos com rigor, não seriam necessários quaisquer stocks de segurança. Como esta situação não acontece, há que manter em stock uma quantidade que proporcione proteção contra uma ou várias eventualidades, o de stock de segurança.

O modelo de revisão contínua quando o PE é demasiado alto, o stock médio aumenta e o custo de posse também. Quando o PE é baixo, o stock médio diminui, mas o risco de rutura aumenta. Quando a quantidade de encomenda é alta, o intervalo de tempo entre encomendas é longo e o custo de posse é elevado. Quando a quantidade encomendada é baixa, o intervalo entre encomendas é curto e o custo de colocação de encomendas é elevado. (Gonçalves, 2010).

De acordo com Schönsleben (2004), a produção ou a compra de um produto tem de ser efetuada, preferencialmente, através de uma previsão de procura. Ou seja, as organizações devem prever o máximo possível da quantidade requerida pelo consumidor, a fim de delinear um planeamento o mais assertivo possível para a sua produção e a gestão de materiais. Slack *et al.* (1997), no âmbito das relações internas da organização, delegam à área de marketing ou vendas a parte da previsão da procura.

Sendo assim, devido à grande incerteza da previsão da procura no planeamento e controlo de stocks, optou-se por apenas se estudar os resultados dos métodos de previsão, pois são uma ferramenta fundamental para diminuir esta incerteza, como vimos anteriormente.

5. Análise dos resultados

Neste capítulo apresenta-se a seleção e análise dos dados, assim como a aplicação dos métodos de previsão já apresentados: Croston, ETS, HW e ARIMA. De seguida, indica-se o procedimento adotado para a previsão da procura e o seu valor para um ano típico. Por fim, apresenta-se uma conclusão onde se faz a análise da previsão obtida.

Para cada método, foi analisado um período com início em janeiro de 2010 e término em agosto de 2016 para uma amostra de 90 artigos. Os erros associados à previsão pelos quatro métodos encontram-se representados na Figuras 14, 15 e 16.

Inicialmente procedeu-se à análise descritiva de uma só série passando posteriormente à análise global dos 90 artigos em estudo.

5.1 Análise descritiva das séries

De modo a poder selecionar a melhor metodologia a adotar, é necessário analisar vários métodos, verificando aqueles que melhor se adaptam aos dados existentes.

Na Figura 11, temos (a exemplo de um dos artigos em estudo) a série original do artigo “filtro de óleo”, com os seus valores expressos por unidades de consumo. Da observação do comportamento da série, nota-se uma certa tendência de decrescimento, o que poderá indicar não-estacionariedade na média.

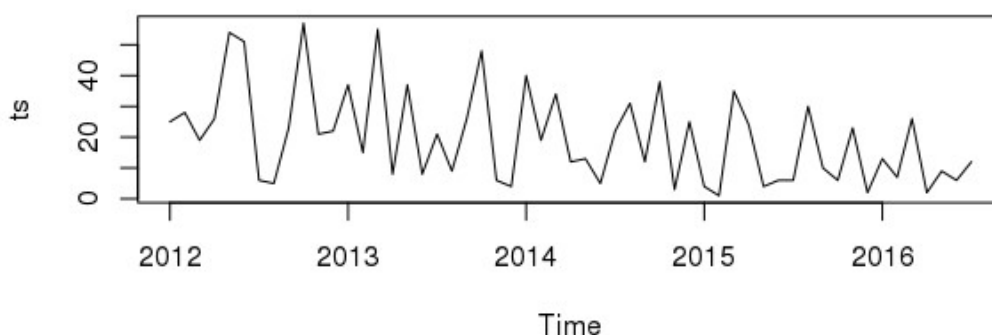


Figura 11 - Série temporal “Filtro Óleo”

De forma a perceber melhor a série temporal foi feita uma decomposição desta, com recurso à função STL do *software* R. Esta função decompõe a série em duas principais componentes: tendência e sazonalidade, podendo depois apresentar os resultados graficamente. Para além destas duas componentes a função apresenta ainda o valor

observado (*observed*) e o que resta da série após ter sido retirada a tendência e a sazonalidade (*seasonal*).

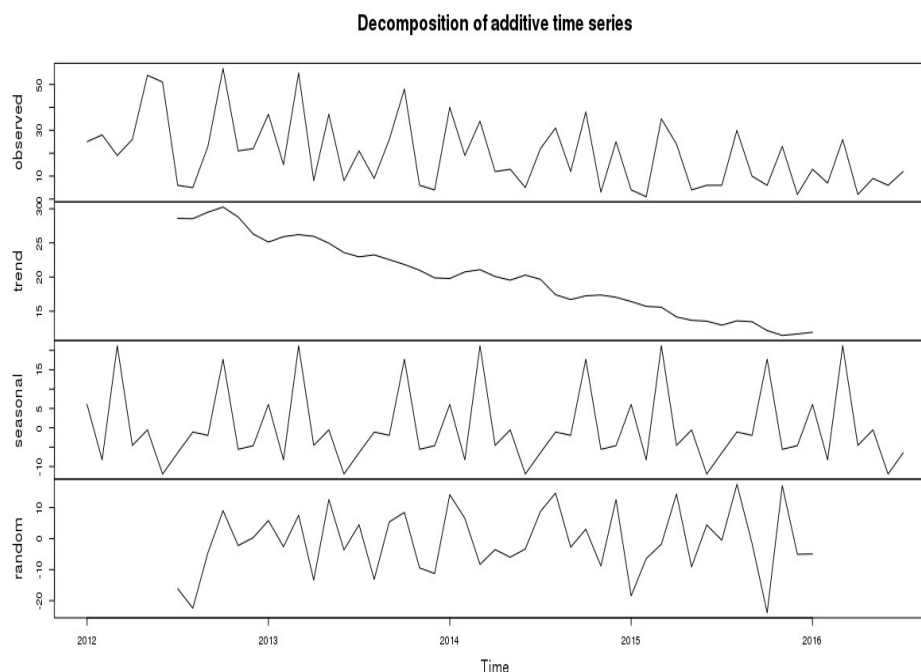


Figura 12 - Decomposição da série temporal “Filtro óleo”

Conforme podemos observar, através da análise da Figura 12, a série apresenta sazonalidade que se repete a cada doze períodos (meses). Quanto à tendência, podemos verificar que esta encontra-se presente na série com um declive negativo. Com a ajuda da decomposição das séries, é nos possível compreender melhor o comportamento dos dados, facilitando, desta forma, a seleção do método mais adequado. Na figura 13, observa-se o gráfico da Função de Autocorrelação da série.

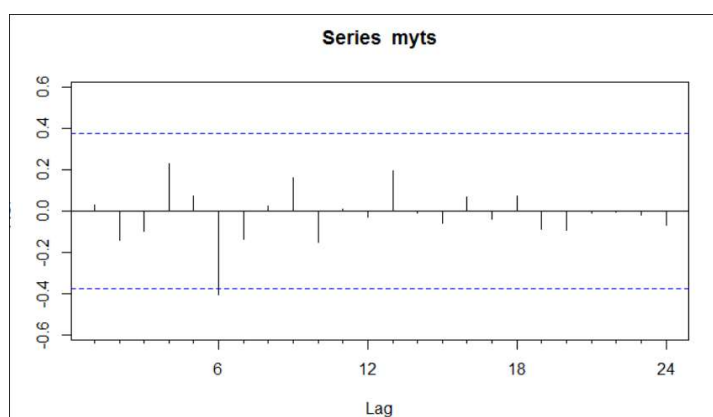


Figura 13 - Gráfico ACF da série original “filtro de óleo”

Uma vez que não há evidência de valores a aproximarem-se dos limites do intervalo de confiança, ou seja, as correlações são todas próximas de zero, conclui-se serem não significativas. Para se confirmar se a série é ou não estacionária procedeu-se ao teste da estacionariedade da série temporal “Filtro óleo” através da função `adf.test{tseries}` no *RStudio* que permite realizar os testes de Dickey-Fuller Aumentado.

Obtendo-se o seguinte *output*:

```
Augmented Dickey-Fuller Test
data: ts
Dickey-Fuller = -3.8089, Lag order = 2, p-
value = 0.0347
alternative hypothesis: stationary
```

Com as hipóteses definidas pelo teste H_0 : série não é estacionária e H_1 : a série é estacionária. Logo sendo $p\text{-value}=0,0347 < 0,05$; rejeitar-se-ia H_0 para $\alpha=0,05$, assim como para qualquer nível de significância superior a 3,5%, o que levaria à rejeição da não estacionariedade da série. Ou seja, apesar da tendência e sazonalidade sugeridas na decomposição da série, tal não é confirmado pela análise do ACF nem pelo teste à estacionariedade para estes níveis.

Assim, depois de se eleger um modelo, e de este ser ajustado aos dados, sendo neste processo estimados os seus parâmetros usando funções do R, foram calculados e comparados os erros MSE, MAE e MAPE (ver Secção 3.6), que são apresentados nas figuras da secção seguinte, com o objetivo de ser verificada a adequação do modelo.

5.2 Comparação entre os erros dos vários métodos

Nesta secção são analisados os resultados obtidos quando ajustados os 4 modelos apresentados nas secções 3.4, 3.5.1, 3.5.2 e 3.5.4 às 90 referências em estudo. Todos os modelos foram avaliados nas 90 séries compostas por 80 observações mensais. Os erros MSE, MAE e MAPE definidos na Secção 3.6 foram seguidamente calculados para cada uma das 90 referências e usando cada um dos 4 métodos e os resultados estão apresentados nas figuras seguintes.

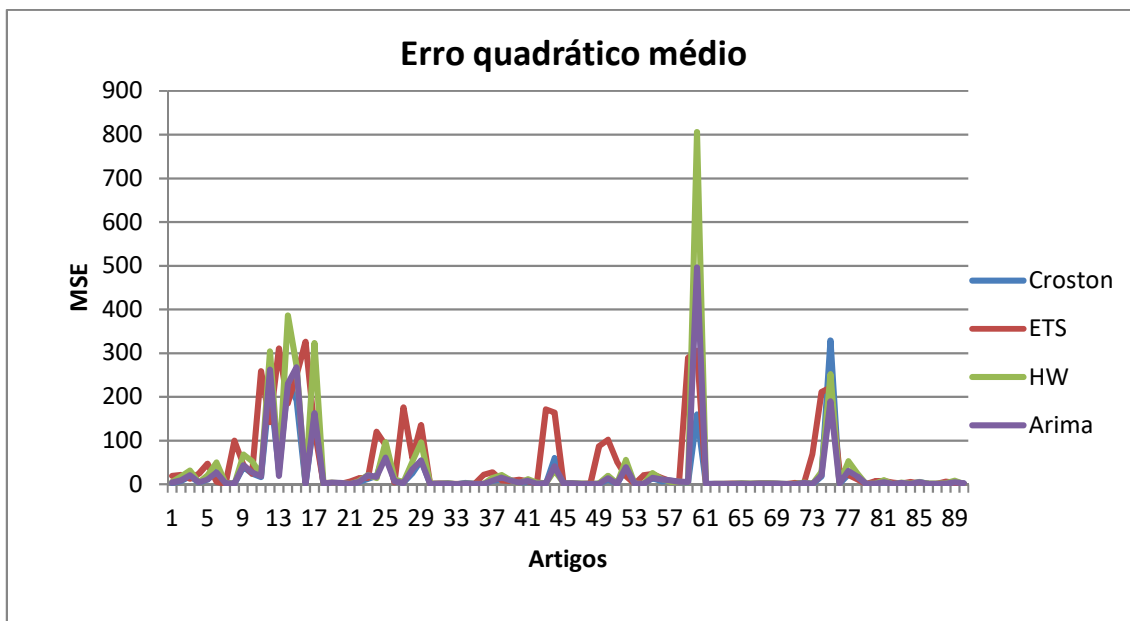


Figura 14 - Erros MSE das 90 referências em estudo

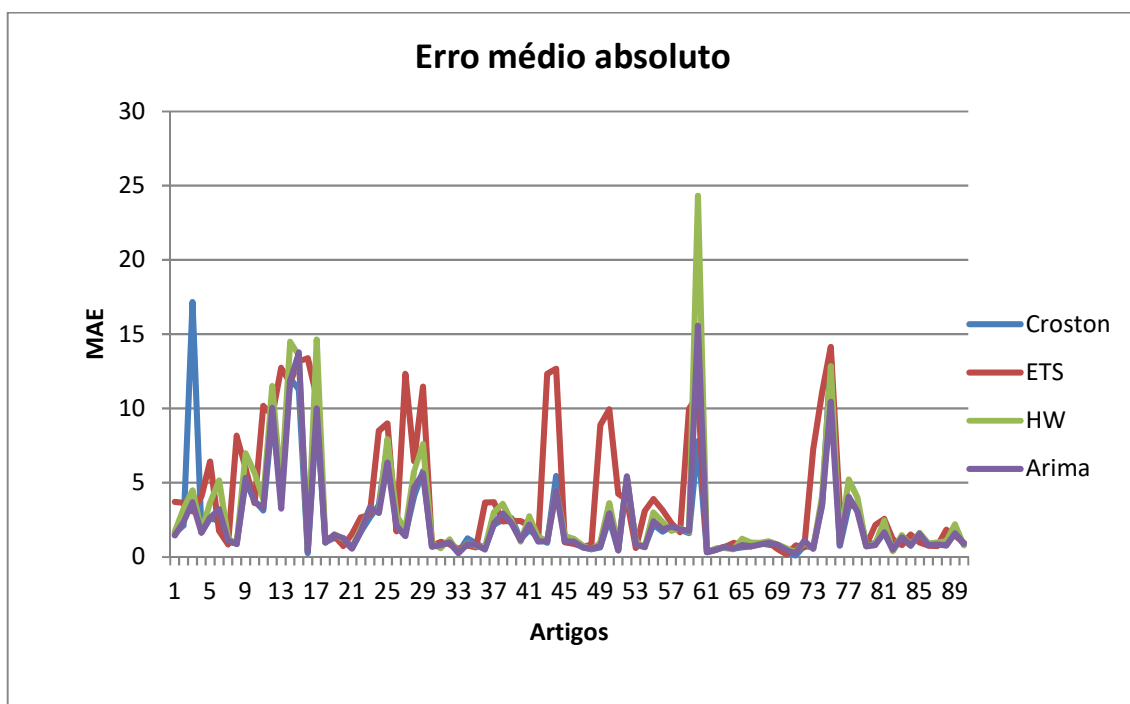


Figura 15 - Erros MAE das 90 referências em estudo

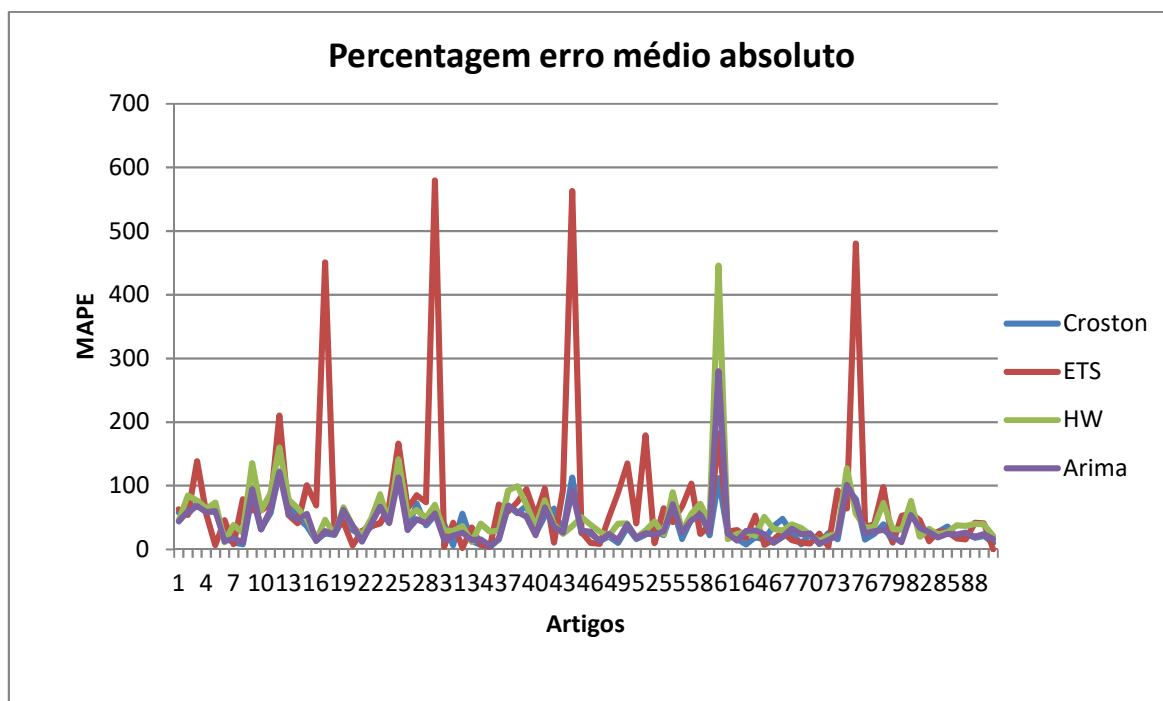


Figura 16 - Erros MAPE das 90 referências em estudo

Da análise das Figuras 14, 15 e 16 verifica-se que a variação dos erros é muito grande devido sobretudo à diferença da ordem de grandeza das encomendas, o que sugere que o MAPE, sendo uma medida de erro relativa, seja a medida mais indicada. O elevado número de artigos e métodos em estudo define algumas limitações na análise visual dos erros. Assim, as Tabelas 1, 2 e 3 apresentam os parâmetros estatísticos descritivos de cada erro, obtidos a partir do *software* SPSS.

Tabela 2 – Erro quadrático médio (MSE)

	N	Minimum	Maximum	Mean	Std. Deviation
cr_MSE	90	,13	329,36	22,2225	55,76298
ets_MSE	90	,20	326,20	50,0289	83,91543
hw_MSE	90	,30	806,39	37,5399	108,75583
arima_MSE	90	,33	495,85	25,8326	72,00321
Valid N (listwise)	90				

Pela análise dos parâmetros estatísticos descritivos do erro quadrático médio, o método Croston produz um erro médio (22,2225) e desvio padrão (55,769298) inferior aos restantes métodos.

Tabela 3 – Erro médio absoluto (MAE)

	N	Minimum	Maximum	Mean	Std. Deviation
cr_MAPE	90	4,06	145,18	38,4639	27,87735
ets_MAPE	90	,00	579,61	71,9422	106,13628
hw_MAPE	90	11,13	445,89	51,8300	51,20152
arima_MAPE	90	5,94	279,70	39,2769	35,02927
Valid N (listwise)	90				

Pela análise dos parâmetros estatísticos descritivos do erro médio absoluto, o método Croston apresenta o menor erro médio (38,4639) e o menor desvio padrão (27,87735).

Tabela 4 – Percentagem Erro Médio Absoluto (MAPE)

	N	Minimum	Maximum	Mean	Std. Deviation
cr_MAE	90	,10	17,17	2,4605	3,10952
ets_MAE	90	,17	14,14	4,0345	4,11026
hw_MAE	90	,27	24,32	2,9327	3,88204
arima_MAE	90	,25	15,58	2,4051	2,95014
Valid N (listwise)	90				

Pela análise dos parâmetros estatísticos descritivos do erro médio absoluto em percentagem, o menor erro médio (2,4051) e o menos desvio padrão (2,95014) é alcançado pelo método ARIMA.

Do ponto vista geral dos erros obtidos, verifica-se que a amplitude dos erros é elevada e o desvio padrão com ordens de grandeza superior ao valor médio do erro. Desta forma, a seleção do método mais adequado não se deverá basear apenas nos valores médios dos erros.

5.3 Teste para comparação das médias dos erros dos diferentes métodos

Como foi dito anteriormente, o responsável pela previsão deve escolher, dentro dos diferentes métodos de previsão, aquele que lhe forneça uma visão do futuro mais realista. Adotou-se a medida de erro MAPE por ser uma medida relativa e que não depende da escala de dados.

Procedeu-se assim ao teste de Friedman, teste não-paramétrico para verificar se a há diferenças significativas na distribuição dos erros para os diferentes métodos, visto a amostra conjunta ser formada por 4 grupos de dados com medidas repetidas. Em particular, testa-se aqui se as medianas dos erros são iguais para todos os métodos ou em alternativa se há pelo menos duas medianas que diferem. O método de Friedman assume que as distribuições das várias variáveis em teste são simétricas, propriedade que implica a igualdade da média, da mediana e da moda. (ver e.g. Teles *et al*, 2016). Desta forma, a comparação dos erros é feita através das suas medianas e não das médias (como nos testes paramétricos), o que, no entanto, é equivalente dado a igualdade entre elas.

```
> friedman.test(dados, groups=metodos, blocks=refs)

Friedman rank sum test

data:  dados, metodos and refs
Friedman chi-squared = 36.433, df = 3, p-value = 6.065e-08
```

Pelo teste acima descrito e como o *p-value* é menor que 0,05, rejeitamos H_0 , logo há pelo menos dois métodos que diferem em termos do erro MAPE. E por este motivo procedemos então ao método das comparações múltiplas, afim de aferir quais as medianas dos erros que diferem entre si. Assim, é usado o consequente teste de comparações múltiplas

emparelhadas descrito em Conover (1999), que no software R é conduzido com a função *posthoc.friedman.conover.test*.

```
> posthoc.friedman.conover.test(y=dados, p.adjust="none",  
groups=metodos, blocks=refs)  
  
Pairwise comparisons using Conover's test for a two-way  
balanced complete block design  
  
data: dados , metodos and refs  
  
      ARIMA    CR      ETS  
CR  0.015    -      -  
ETS < 2e-16 < 2e-16 -  
HW  < 2e-16 < 2e-16 3.2e-06
```

Pelo referido teste, concluímos que os *p-values* associados a cada par são inferiores a 0,05, o que nos leva a concluir, para esse nível de significância, que os métodos são todos diferentes entre si, existindo um método que é melhor do que o outro¹¹ (ou seja, o que tiver o erro MAPE menor e consequentemente a sua mediana).

Olhando agora para os dados amostrais, em particular para as médias dos erros MAPE, e restante análise dos parâmetros estatísticos descritivos, o método ARIMA é o tem menor erro médio e o menor desvio padrão. Conclui-se, portanto, que será este o método com menores erros MAPE e logo será o modelo a adotar.

Alternativamente, e como os métodos Arima e Croston foram os que apresentaram os erros MAPE mais baixos (ver Tabela 4), foi efetuado um *t-test* para amostras emparelhadas, teste paramétrico, a fim de serem comparadas as médias dos erros MAPE entre esses dois métodos. Note-se que, tendo em conta que os erros são variáveis contínuas das quais temos 90 observações de cada (dimensão que permite aproximações à normal através do Teorema do Limite Central¹²), o *t-test* poderá ser aqui aplicado e implementado através da função *t-test* do *software* R.

¹¹ As medianas diferem, logo uma é maior do que a outra.

¹² O número de observações é superior 30.


```

> t.test(ARIMA,CROSTON,mu=0,alternative="less",paired=TRUE)

Paired t-test

data:  ARIMA and CROSTON

t = 0.36042, df = 89, p-value = 0.6403

alternative hypothesis: true difference in means is less than 0

> summary(CROSTON)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
4.062  18.800  28.400  38.460  55.930 145.200

> summary(ARIMA)

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5.938  20.700  28.260  39.280  53.430 279.700

```

Pela análise do *p-value* não se rejeita H_0 , logo não pode concluir que as médias dos erros sejam diferentes para os dois métodos. Sendo neste caso o teste é inclusivo no sentido de se concluir qual dos métodos será o melhor.

Alternativamente, poder-se-ia usar aqui uma análise de variância (ANOVA) a um fator com medidas repetidas, para verificar se os erros MAPE associados aos quatro métodos são significativamente diferentes. Apesar deste teste ser o alternativo paramétrico correspondente ao teste de Friedman, ele não é muito corrente e a sua implementação no R ainda não está muito desenvolvida, podendo ser efetuado através da função “Anova” do *package car* do *software* R, mas não existindo até ao momento a implementação das correspondentes comparações múltiplas necessárias quando há rejeição da hipótese de todas as médias serem iguais.

Os resultados encontram-se em seguida:

```

> Anova(metodos.model, idata=dados, idesign=~metodos.frame)

Anova Table (Type III tests)

Response: dados

              Sum Sq Df F value    Pr(>F)
(Intercept) 3150.4   1  244.05 < 2.2e-16 ***
Residuals   4634.1 359

```

Da análise dos resultados do teste ANOVA, vê-se que, sendo o *p-value* pequeno (inferior a 0,05) se pode concluir pela rejeição de H_0 , o que implica que os métodos diferem em

termos de erros médios. No entanto, para se poder concluir qual o melhor método em termos de produzir o menor erro MAPE médio, dever-se-ia proceder ao cálculo das comparações múltiplas, função que até ao momento não se encontra disponível no R. Logo podemos apenas concluir que o ARIMA é o melhor método usando o teste de Friedman e respetivas comparações múltiplas.

5.4 Exemplos da aplicação dos métodos

Como vimos anteriormente os modelos com melhor resultado foram o Croston e o modelo ARIMA, segue-se assim a previsão para 2 séries temporais, uma série intermitente e outra com observações de valores elevados.

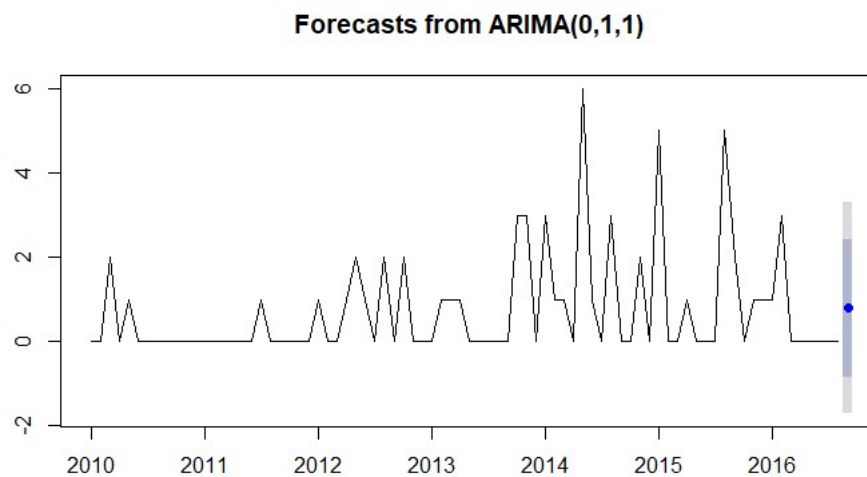


Figura 17 – Série intermitente, espigão de plástico – modelo ARIMA (0,1,1)

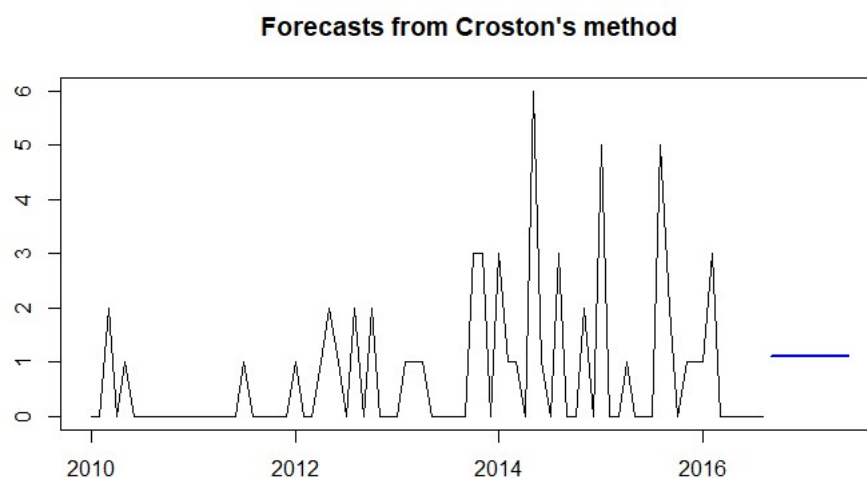


Figura 18 – Série intermitente, espigão de plástico – Método de Croston

Nas figuras 17 e 18 está representada a série intermitente e respetiva previsão, usando o modelo ARIMA (figura 17) e o modelo Croston (figura 18).

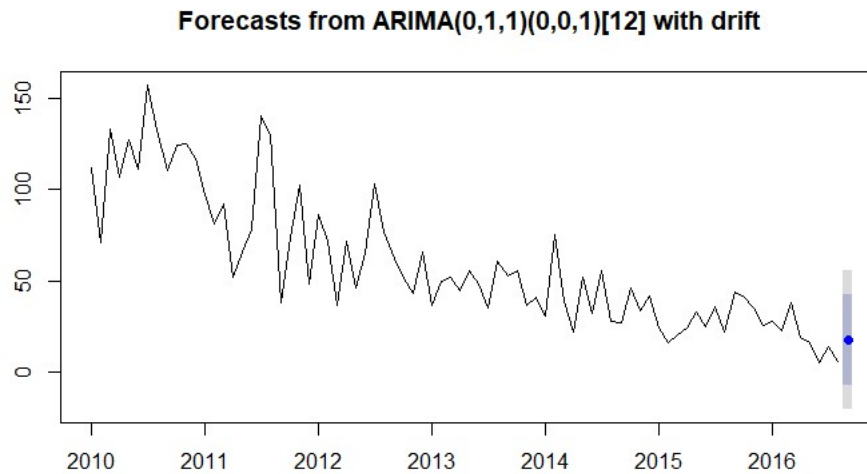


Figura 19 – Amortecedor – modelo ARIMA (0,1,1)

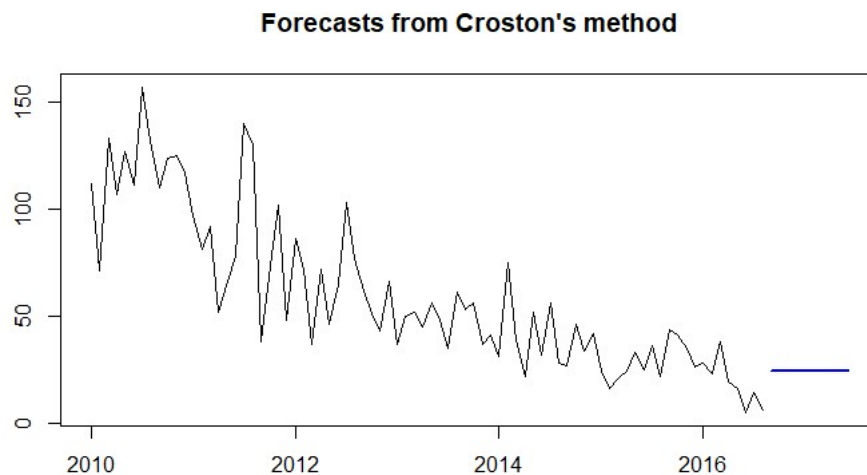


Figura 20 - Amortecedor – Método de Croston

Nas figuras 19 e 20 está representada a série com observações de valores elevados e respetiva previsão, usando o modelo ARIMA (figura 19) e o modelo Croston (figura 20). Nas previsões ARIMA a cinzento escuro está representado o intervalo de confiança para previsão a 95% e a cinzento claro a 99%.

Na primeira série é de esperar que para a primeira série que a previsão que vá mais de acordo com a realidade seja obtida pelo método de Croston por a série ser intermitente.

6. Conclusões

A crescente procura por níveis cada vez mais elevados de serviço e de qualidade pressiona as organizações a terem maior agilidade e a direcionarem as suas atividades para a satisfação do cliente final, delineando todos os seus sistemas e procedimentos com o objetivo principal de melhorar a velocidade e a confiança da resposta. A existência de um sistema de gestão de stocks eficiente e otimizado é uma componente essencial para a obtenção de níveis elevados de serviço.

A previsão da procura assume elevada importância na gestão de stocks de peças sobressalentes. Devido à procura deste tipo de artigos apresentar sazonalidade e intermitência a previsão da procura torna-se uma tarefa difícil e complexa.

A literatura existente neste domínio, em particular no caso da procura irregular, não é extensa, referindo apenas a criação de um nível de stock seguro para lidar com eventuais ruturas, conforme é referido por Chen *et al.* (2009).

Nesta dissertação foram analisados vários modelos para previsão da procura e foram utilizadas as medidas de erro MSE, MAE e MAPE de forma a avaliar o desempenho dos modelos. Para o conjunto de teste o método ARIMA foi o que conseguiu melhores resultados de acordo com a medida de erro MAPE.

Sendo que o objetivo final da M Coutinho Peças é a satisfação do cliente, o nível de stock deve ser o mais próximo possível da procura desejada de forma a que o cliente seja servido no menor tempo possível.

O método que mais se aproximou deste objetivo foi o método ARIMA, pelo teste de Friedman, tornando-se assim no método adotar para a implementação de um sistema de gestão de stocks para a M Coutinho Peças. Em alternativa o método será considerado em séries temporais que se relevaram com procura intermitente.

Tópicos a Desenvolver no Futuro

Tendo em conta uma reflexão a posteriori dos temas desenvolvidos, assim como, os aspetos metodológicos expostos no decorrer do trabalho, surge a necessidade de ponderar sobre algumas questões específicas.

O emprego de modelos não causais, em vez de modelos causais, requer, desde logo, esta reflexão. Uma vez que as séries em estudo são relativas a vendas, o uso de modelos causais com implementação de variáveis de marketing como variáveis explicativas poderá apresentar melhores resultados das previsões, assim como permitir realizar uma análise de cenários associados, nomeadamente, a variações de preços de venda. A previsão de variações de preços, promoções ou gastos em publicidade poderá permitir antecipar as vendas de uma forma mais precisa (Kioulafas, 1985).

7. Referências bibliográficas

- Ballou, R. H. (2004). *Business logistics and supply chain management*. Pearson Prentice Hall.
- Bartezzaghi E., Verganti R., Zotteri G., (1999). A simulation framework for forecasting uncertain lumpy demand, *International Journal of Production Economics*, Vol. 59
- Box, G. P. E., Jenkins, G. M. e Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control*, 3rd Edition, Prentice-Hall.
- Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods*, 2nd Edition, New York: Springer-Verlag.
- Brown, R. (1959). *Statistical Forecasting for Inventory Control*, McGraw-Hill, New York.
- Brown, R. (1963). *Smoothing, Forecasting, and prediction of Discrete Time series*, Prentice-Hall, New Jersey.
- Caiado, J. (2002). *Métodos de previsão em Economia e Gestão: Desenvolvimentos recentes*, Inforbanca.
- Cavalieri, S., Garetti, M., Macchi, M. e Pinto, R. (2008). A decision-making framework for managing maintenance spare parts.
- Chen F.L., Chen Y.C. (2009). *An investigation of forecasting critical spare parts requirement*, Em World congress on computer science and information engineering, p. 225-230.
- Conover, W. J. (1999). *Practical Nonparametric Statistics*, 3rd Edition. New York: John Wiley & Sons.

- Cowpertwait P., Metcalfe A. (2009). *Introductory Time Series with R*, Springer New York.
- Croston, J. D. (1972). *Forecasting and Stock Control for Intermittent Demand*, Operational Research Quarterly, v.23-3, p. 289-303
- Davydenko, A. e Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts, *International Journal of Forecasting*, Vol. 29, N° 3, p. 510–522.
- Enders, W. (2009), *Applied Econometric Time Series*, Wiley Series in Probability and Statistics, Wiley
- Fernandes, R., B. Gouveia e C. Pinho (2013), Integrated inventory valuation in multi-echelon production/distribution systems, *International Journal of Production Research*, Vol. 51, N° 9, p. 2578-2592.
- Fernie, J. e L. Sparks (2004). *Retail logistics: Changes and challenges*, in *Logistics and Retail Management*, The Institute of Logistics and Transport, 2nd Edition, pp. 1-25, Great Britain.
- Franceschini, F., Galetto, M., & Maisano, D. (2007). *Management by Measurement: Designing Key Indicators and Performance Measurement Systems*, Springer Verlag.
- Gaither, N., Fraizer, G. (2005). *Administração da produção e Operações*, Pioneira Thompson Learning, 8ª edição, São Paulo.
- Gamberini, R., Lolli, F., Rimini, B. e Sgarbossa, F. (2000). Forecasting of Sporadic Comparison between Holt-Winters and SARIMA Methods, *Journal of Forecasting*.
- Gardner, E. S. (2006). Exponential smoothing: The state of the art - Part II, *Journal of Forecasting*.

- Gonçalves, J. (2010). *Gestão de Aprovisionamento – Stocks, Previsão, Compras*, Publindústria, Edições Técnicas.
- Gujarati DN (2004). *Basic Econometrics*, McGraw Hill.
- Heizer, J., Render, B. (2000). *Operations Management*, 6ªEd, Prentice Hall.
- Holt, C. (1957). *Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages*, ONR Research Memorandum, Carnegie Institute of Technology.
- Hyndman R., Athanasopoulos G. (2012). *Forecasting: Principles and practice*.
- Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. In *Journal of Statistical Software*, Volume 27.
- K. E. Kioulafas (1985). An Application of Multiple Regression Analysis to the Greek Beer Market, *The Journal of the Operational Research Society*.
- Kwiatkowski D, Phillips PC, Schmidt P, Shin Y (1992). Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*.
- M. Zied Babai, Mohammad. M. (2011). *Impact of temporal aggregation on stock control performance of intermittent*, Omega, 9.
- Makridakis, S., Wheelwright, S. C e Hyndman, R. J. (1998). *Forecasting Methods and Applications*, 3ª edição, New York: John Wiley & Sons.
- Murteira, B., Muller, D. e Turkman, K. (2000). *Análise de sucessões cronológicas*, McGrawHill.
- Nenes, G., S. Panagiotidou e G. Tagaras (2010). Inventory management of multiple items with irregular demand: A case study, *European Journal of Operational Research*, Vol. 205, Nº 2, pp. 313-324.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

- Richard B. Chase, F. R. (2006). *Administração da Produção e Operações para Vantagens Competitivas*, 11th Edition, Mc Graw Hill.
- Teles P., Ramos A., Figueiredo A., Figueiredo F. (2016). *Estatística Descritiva e Probalidades: Problemas resolvidos e propostos em R*, Escolar Editora, p. 536-538.
- Tratar, F. L. (2010). Joint optimisation of demand forecasting and stock control parameters, *International Journal of Production Economics*, vol. 127, p. 173-179.
- Wei, W. W. S. (2007). *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed., Addison-Wesley Publishing Company, Redwood City, California.
- Winters, P. (1960). *Forecasting sales by exponentially weighted moving averages*. *Management Science*, p. 324–342.

8. Anexos

Anexo A. Previsão utilizando o método do Alisamento Exponencial Simples

```
library(RMySQL)
library(forecast)
library(lubridate)
source("util.R")

generate_graphs <- FALSE
fill_forecast_data <- TRUE
debug <- TRUE
contador = 1
cont = 80
prev = matrix(9999,100,80)

if (isTRUE(debug)) { print("Connecting MySQL") }
drv = dbDriver("MySQL")
con =
dbConnect(drv,host="lpvs1.internal",port=,dbname="MCPC_SI",user="jppai
s",pass="")
if (isTRUE(debug)) { print("Fetching itens to forecast") }
ds = dbSendQuery(con,statement="
        SELECT
        c.Armazem,
        c.Marca,
        c.Referencia
        FROM
        tbl_stock_controlo c
        WHERE
        c.Armazem IN ('71','78','471','478')
        AND c.Marca = '17'
        -- AND Referencia = '1109AP'
        ORDER BY
        c.Armazem,
        c.Marca,
        c.Referencia
        LIMIT 100
        ")
data <- fetch(ds, n = -1)
for(i in 1:100) {
  Armazem = data$Armazem[i]
  Marca = data$Marca[i]
  Referencia = data$Referencia[i]
  ReferenciaCompleta =
paste(Marca,paste("/",paste(Referencia,sep=""),sep=""),sep="")
  if (isTRUE(debug)) {
    print(
      paste(
        paste(
          paste(
            paste(
              paste(
                i,
                "/",

```

```

        sep = ""
    ),
    (dim(data)-1)[1],
    sep = ""
),
" - ",
sep = ""
),
Armazem,
sep = ""
),
" - ",
sep = ""
),
ReferenciaCompleta,
sep = ""
)
)
}
SQL = paste("
    SELECT
    SUM(IFNULL(det.Qtd,0)) AS Qtd
    FROM
    MCBT_DW.Datas dat
    LEFT JOIN MCPC_SI.tbl_facturas_detalhe det ON
    dat.Date = det.`Data`
    AND det.CodCliente NOT IN
('K0071','K0078','K0471','K0478')
    AND det.`Data` >= '2010-01-01'
    AND det.Referencia = '',
    paste(ReferenciaCompleta,
    paste('
        AND det.Armazem = '',
        paste(Armazem,
        paste('
            WHERE
            dat.`Date` >= '2010-01-01'
            AND dat.`Date` <
CAST(CONCAT (DATE_FORMAT (CURDATE(), '%Y-%m'), '-01') AS DATE)
            AND dat.WeekDay < 5
            GROUP BY
            YEAR(dat.Date),
            MONTH(dat.Date)

'', sep=""), sep=""), sep=""), sep=""), sep="")
    )
SQL = gsub("\n", "", SQL)
dsl = dbGetQuery(con, statement=SQL)

for (a in c(1,12,24,36,48,60,72)) {
    ano=c(year(now()-months(a)))

    if (ano[1]==2016) {
        fim=c(1:8)}
    if(ano[1]!=2016){
        fim=c(-3,-2,-1,0,1,2,3,4,5,6,7,8)}

    for (j in fim) {

```

```

end=c(year(now()-months(a)), month(now()-months(j)))
#print(end)
ts = ts(dsl, start=c(2010, 1), end, frequency=12)
#print(ts)

try(fit_ets <- ets(ts))

if (exists("fit_ets")) {
  fc = forecast(fit_ets,1)
}

# Goes on
if (exists("fc")) {
  previsao = max(c(0,round(fc$mean[1], digits = 0)))
  prev[contador,cont]=previsao
  rm(fc)

}
cont = cont-1
if (cont==4){
  break}
#plot(fc,main=paste(Armazem,paste(" -
",ReferenciaCompleta,sep=""), sep=""))
}
}
# print(prev)
contador = contador+1
cont=80
}

write.table(prev, "ets.txt", sep="\t", row.names = FALSE, col.names =
FALSE)

dbDisconnect(con)
if (isTRUE(debug)) { print("Done") }

```

Anexo B. Previsão utilizando o método Holt-Winters

```
library(RMySQL)
library(forecast)
library(lubridate)
source("util.R")

generate_graphs      <- FALSE
fill_forecast_data   <- TRUE
debug                <- TRUE
contador              = 1
cont = 80
prev = matrix(9999,100,80)

if (isTRUE(debug)) { print("Connecting MySQL") }
drv = dbDriver("MySQL")
con =
dbConnect(drv,host="lpvsl.internal",port=,dbname="MCPC_SI",user="jppai
s",pass="")
if (isTRUE(debug)) { print("Fetching itens to forecast") }
ds = dbSendQuery(con,statement="
SELECT
  c.Armazem,
  c.Marca,
  c.Referencia
FROM
  tbl_stock_controlo c
WHERE
  c.Armazem IN ('71','78','471','478')
AND c.Marca = '17'
-- AND Referencia = '1109AP'
ORDER BY
  c.Armazem,
  c.Marca,
  c.Referencia
LIMIT 100
")
data <- fetch(ds, n = -1)
for(i in 1:100) {
  Armazem = data$Armazem[i]
  Marca = data$Marca[i]
  Referencia = data$Referencia[i]
  ReferenciaCompleta =
paste(Marca,paste("/",paste(Referencia,sep=""),sep=""),sep="")
  if (isTRUE(debug)) {
    print(
      paste(
        paste(
          paste(
            paste(
              paste(
                i,
                "/",
                sep = ""
              ),
            ),
          ),
        ),
      ),
    ),
  }
```

```

        (dim(data)-1)[1],
        sep = ""
    ),
    " - ",
    sep = ""
),
    Armazem,
    sep = ""
),
    " - ",
    sep = ""
),
    ReferenciaCompleta,
    sep = ""
)
}
SQL = paste("
    SELECT
    SUM(IFNULL(det.Qtd,0)) AS Qtd
    FROM
    MCBT_DW.Datas dat
    LEFT JOIN MCPC_SI.tbl_facturas_detalhe det ON
    dat.Date = det.`Data`
    AND det.CodCliente NOT IN
('K0071','K0078','K0471','K0478')
    AND det.`Data` >= '2010-01-01'
    AND det.Referencia = '',
    paste(ReferenciaCompleta,
    paste("
        AND det.Armazem = '',
        paste(Armazem,
        paste("
            WHERE
            dat.`Date` >= '2010-01-01'
            AND dat.`Date` <
CAST(CONCAT(
DATE_FORMAT(CURDATE(), '%Y-%m'), '-01') AS DATE)
            AND dat.WeekDay < 5
            GROUP BY
            YEAR(dat.Date),
            MONTH(dat.Date)

", sep=""), sep=""), sep=""), sep=""), sep="")
)
SQL = gsub("\n", "", SQL)
dsl = dbGetQuery(con, statement=SQL)

for (a in c(1,12,24,36,48,60,72)) {
    ano=c(year(now()-months(a)))

    if (ano[1]==2016) {
        fim=c(1:8)}
    if(ano[1]!=2016){
        fim=c(-3,-2,-1,0,1,2,3,4,5,6,7,8)}

    for (j in fim) {

        end=c(year(now()-months(a)), month(now()-months(j)))
        #print(end)
    }
}

```

```

ts = ts(dsl, start=c(2010, 1), end, frequency=12)

try(fit_hw <- HoltWinters(ts,seasonal = c("additive"))))

if (exists("fit_hw")) {
  fc_hw = forecast(fit_hw,1)
  fc <- fc_hw
  max_forecast = fc_hw$mean[1]
}

# Goes on
if (exists("fc")) {
  previsao = max(c(0,round(fc$mean[1], digits = 0)))
  prev[contador,cont]=previsao
  rm(fc)
  #print(previsao)
  #print(cont)

}
cont = cont-1

#plot(fc,main=paste(Armazem,paste(" -
",ReferenciaCompleta,sep=""), sep=""))
}
if (cont==24){
  break}
}
# print(prev)
contador = contador+1
cont=80

}

write.table(prev, "holt.txt", sep="\t", row.names = FALSE, col.names =
FALSE)

dbDisconnect(con)
if (isTRUE(debug)) { print("Done") }

```

Anexo C. Previsão utilizando o método ARIMA

```
library(RMySQL)
library(forecast)
library(lubridate)
source("util.R")

generate_graphs      <- FALSE
fill_forecast_data   <- TRUE
debug                <- TRUE
contador              = 1
cont = 80
prev = matrix(9999,100,80)

if (isTRUE(debug)) { print("Connecting MySQL") }
drv = dbDriver("MySQL")
con =
dbConnect(drv,host="lpvsl.internal",port=,dbname="MCPC_SI",user="jppai
s",pass="")
if (isTRUE(debug)) { print("Fetching itens to forecast") }
ds = dbSendQuery(con,statement="
SELECT
  c.Armazem,
  c.Marca,
  c.Referencia
FROM
  tbl_stock_controlo c
WHERE
  c.Armazem IN ('71','78','471','478')
AND c.Marca = '17'
-- AND Referencia = '1109AP'
ORDER BY
  c.Armazem,
  c.Marca,
  c.Referencia
LIMIT 100
")
data <- fetch(ds, n = -1)
for(i in 1:100) {
  Armazem = data$Armazem[i]
  Marca = data$Marca[i]
  Referencia = data$Referencia[i]
  ReferenciaCompleta =
paste(Marca,paste("/",paste(Referencia,sep=""),sep=""),sep="")
  if (isTRUE(debug)) {
    print(
      paste(
        paste(
          paste(
            paste(
              i,
              "/",
              sep = ""
            ),
          ),
        ),
      ),
    ),
  }
```



```

        (dim(data)-1)[1],
        sep = ""
    ),
    " - ",
    sep = ""
),
    Armazem,
    sep = ""
),
    " - ",
    sep = ""
),
    ReferenciaCompleta,
    sep = ""
)
}
SQL = paste("
    SELECT
    SUM(IFNULL(det.Qtd,0)) AS Qtd
    FROM
    MCBT_DW.Datas dat
    LEFT JOIN MCPC_SI.tbl_facturas_detalhe det ON
    dat.Date = det.`Data`
    AND det.CodCliente NOT IN
('K0071','K0078','K0471','K0478')
    AND det.`Data` >= '2010-01-01'
    AND det.Referencia = '',
    paste(ReferenciaCompleta,
    paste("
        AND det.Armazem = '',
        paste(Armazem,
        paste("
            WHERE
            dat.`Date` >= '2010-01-01'
            AND dat.`Date` <
CAST(CONCAT(
DATE_FORMAT(CURDATE(), '%Y-%m'), '-01') AS DATE)
            AND dat.WeekDay < 5
            GROUP BY
            YEAR(dat.Date),
            MONTH(dat.Date)

", sep=""), sep=""), sep=""), sep=""), sep="")
)
SQL = gsub("\n", "", SQL)
dsl = dbGetQuery(con, statement=SQL)

for (a in c(1,12,24,36,48,60,72)) {
    ano=c(year(now()-months(a)))

    if (ano[1]==2016) {
        fim=c(1:8)}
    if(ano[1]!=2016){
        fim=c(-3,-2,-1,0,1,2,3,4,5,6,7,8)}

    for (j in fim) {

        end=c(year(now()-months(a)), month(now()-months(j)))
        #print(end)
    }
}

```

```

ts = ts(ds1, start=c(2010, 1), end, frequency=12)
#print(ds1)

try(fit_ar <- auto.arima(ts))

if (exists("fit_ar")) {
  fc_ar = forecast(fit_ar,1)
  fc <- fc_ar
  max_forecast = fc_ar$mean[1]
}

# Goes on
if (exists("fc")) {
  previsao = max(c(0,round(fc$mean[1], digits = 0)))
  prev[contador,cont]=previsao
  rm(fc)
  # print(previsao)

}
cont = cont-1
if (cont==0){
  break}
#plot(fc,main=paste(Armazem,paste(" -
",ReferenciaCompleta,sep=""), sep=""))
}
}
# print(prev)
contador = contador+1
cont=80
}

write.table(prev, "autoarima.txt", sep="\t", row.names = FALSE, col.names
= FALSE)

dbDisconnect(con)
if (isTRUE(debug)) { print("Done") }

```

Anexo D. Previsão utilizando o método de Croston

```
library(RMySQL)
library(forecast)
library(lubridate)
source("util.R")

generate_graphs      <- FALSE
fill_forecast_data   <- TRUE
debug                <- TRUE
contador              = 1
cont = 80
prev = matrix(9999,100,80)

if (isTRUE(debug)) { print("Connecting MySQL") }
drv = dbDriver("MySQL")
con =
dbConnect(drv,host="lpvsl.internal",port=,dbname="MCPC_SI",user="jppai
s",pass="")
if (isTRUE(debug)) { print("Fetching itens to forecast") }
ds = dbSendQuery(con,statement="
SELECT
  c.Armazem,
  c.Marca,
  c.Referencia
FROM
  tbl_stock_controlo c
WHERE
  c.Armazem IN ('71','78','471','478')
AND c.Marca = '17'
-- AND Referencia = '1109AP'
ORDER BY
  c.Armazem,
  c.Marca,
  c.Referencia
LIMIT 100
")
data <- fetch(ds, n = -1)
for(i in 1:100) {
  Armazem = data$Armazem[i]
  Marca = data$Marca[i]
  Referencia = data$Referencia[i]
  ReferenciaCompleta =
paste(Marca,paste("/",paste(Referencia,sep=""),sep=""),sep="")
  if (isTRUE(debug)) {
    print(
      paste(
        paste(
          paste(
            paste(
              i,
              "/",
              sep = ""
            ),
          ),
        ),
      ),
    ),
  }
```

```

        (dim(data)-1)[1],
        sep = ""
    ),
    " - ",
    sep = ""
),
    Armazem,
    sep = ""
),
    " - ",
    sep = ""
),
    ReferenciaCompleta,
    sep = ""
)
}
SQL = paste("
    SELECT
    SUM(IFNULL(det.Qtd,0)) AS Qtd
    FROM
    MCBT_DW.Datas dat
    LEFT JOIN MCPC_SI.tbl_facturas_detalhe det ON
    dat.Date = det.`Data`
    AND det.CodCliente NOT IN
('K0071','K0078','K0471','K0478')
    AND det.`Data` >= '2010-01-01'
    AND det.Referencia = '',
    paste(ReferenciaCompleta,
    paste("'
        AND det.Armazem = '',
        paste(Armazem,
        paste("'
            WHERE
            dat.`Date` >= '2010-01-01'
            AND dat.`Date` <
CAST(CONCAT(
DATE_FORMAT(CURDATE(), '%Y-%m'), '-01') AS DATE)
            AND dat.WeekDay < 5
            GROUP BY
            YEAR(dat.Date),
            MONTH(dat.Date)

", sep=""), sep=""), sep=""), sep=""), sep="")
)
SQL = gsub("\n", "", SQL)
dsl = dbGetQuery(con, statement=SQL)

for (a in c(1,12,24,36,48,60,72)) {
    ano=c(year(now()-months(a)))

    if (ano[1]==2016) {
        fim=c(1:8)}
    if(ano[1]!=2016){
        fim=c(-3,-2,-1,0,1,2,3,4,5,6,7,8)}

    for (j in fim) {

        end=c(year(now()-months(a)), month(now()-months(j)))
        #print(end)
    }
}

```

```

ts = ts(ds1, start=c(2010, 1), end, frequency=12)
#print(ds1)

if (sum(ts < 0)[1] <= 0) {
  try(fit_cr <- croston(ts))
}
#options(warn=-1)
#if (all(ts) >= 0) {
#  try(fit_cr <- croston(ts))
#}

if (exists("fit_cr")) {
  fc_cr = forecast(fit_cr,1)
  fc <- fc_cr
}

# Goes on
if (exists("fc")) {
  previsao = max(c(0, round(fc$mean[1], digits = 0)))
  prev[contador, cont] = previsao
  rm(fc)
  # print(previsao)

}
cont = cont-1
if (cont==0){
break}
  #plot(fc, main=paste(Armazem, paste(" -
", ReferenciaCompleta, sep=""), sep=""))
}
}
# print(prev)
contador = contador+1
cont=80
}

write.table(prev, "testar_exportar3.txt", sep="\t", row.names =
FALSE, col.names = FALSE)

dbDisconnect(con)
if (isTRUE(debug)) { print("Done") }

```

Anexo E – Resultados do Teste Wilcoxon para Erro Quadrático Médio

Ranks

		N	Mean Rank	Sum of Ranks
arima_MSE - cr_MSE	Negative Ranks	27 ^a	44,13	1191,50
	Positive Ranks	61 ^b	44,66	2724,50
	Ties	2 ^c		
	Total	90		
arima_MSE - ets_MSE	Negative Ranks	50 ^d	53,30	2665,00
	Positive Ranks	40 ^e	35,75	1430,00
	Ties	0 ^f		
	Total	90		
arima_MSE - hw_MSE	Negative Ranks	72 ^g	48,79	3513,00
	Positive Ranks	17 ^h	28,94	492,00
	Ties	1 ⁱ		
	Total	90		

a. arima_MSE < cr_MSE

b. arima_MSE > cr_MSE

c. arima_MSE = cr_MSE

d. arima_MSE < ets_MSE

e. arima_MSE > ets_MSE

f. arima_MSE = ets_MSE

g. arima_MSE < hw_MSE

h. arima_MSE > hw_MSE

i. arima_MSE = hw_MSE

Test Statistics^c

	arima_MSE - cr_MSE	arima_MSE - ets_MSE	arima_MSE - hw_MSE
Z	-3,189 ^a	-2,485 ^b	-6,180 ^b
Asymp. Sig. (2-tailed)	,001	,013	,000

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
cr_MSE - ets_MSE	Negative Ranks	54 ^a	53,72	2901,00
	Positive Ranks	36 ^b	33,17	1194,00
	Ties	0 ^c		
	Total	90		
cr_MSE - hw_MSE	Negative Ranks	73 ^d	48,10	3511,00
	Positive Ranks	17 ^e	34,35	584,00
	Ties	0 ^f		
	Total	90		
cr_MSE - arima_MSE	Negative Ranks	61 ^g	44,66	2724,50
	Positive Ranks	27 ^h	44,13	1191,50
	Ties	2 ⁱ		
	Total	90		

Test Statistics^c

	arma_MSE - cr_MSE	arma_MSE - ets_MSE	arma_MSE - hw_MSE
Z	-3,189 ^a	-2,485 ^b	-6,180 ^b
Asymp. Sig. (2-tailed)	,001	,013	,000

a. cr_MSE < ets_MSE

b. cr_MSE > ets_MSE

c. cr_MSE = ets_MSE

d. cr_MSE < hw_MSE

e. cr_MSE > hw_MSE

f. cr_MSE = hw_MSE

g. cr_MSE < arma_MSE

h. cr_MSE > arma_MSE

i. cr_MSE = arma_MSE

Test Statistics^b

	cr_MSE - ets_MSE	cr_MSE - hw_MSE	cr_MSE - arma_MSE
Z	-3,434 ^a	-5,889 ^a	-3,189 ^a
Asymp. Sig. (2-tailed)	,001	,000	,001

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Anexo F – Resultados do Teste Wilcoxon para Erro Absoluto Médio

Ranks		N	Mean Rank	Sum of Ranks
arima_MAE - cr_MAE	Negative Ranks	38 ^a	40,30	1531,50
	Positive Ranks	47 ^b	45,18	2123,50
	Ties	5 ^c		
	Total	90		
arima_MAE - ets_MAE	Negative Ranks	56 ^d	54,39	3046,00
	Positive Ranks	34 ^e	30,85	1049,00
	Ties	0 ^f		
	Total	90		
arima_MAE - hw_MAE	Negative Ranks	70 ^g	50,55	3538,50
	Positive Ranks	19 ^h	24,55	466,50
	Ties	1 ⁱ		
	Total	90		

a. arima_MAE < cr_MAE

b. arima_MAE > cr_MAE

c. arima_MAE = cr_MAE

d. arima_MAE < ets_MAE

e. arima_MAE > ets_MAE

f. arima_MAE = ets_MAE

g. arima_MAE < hw_MAE

h. arima_MAE > hw_MAE

i. arima_MAE = hw_MAE

Test Statistics ^c			
	arima_MAE - cr_MAE	arima_MAE - ets_MAE	arima_MAE - hw_MAE
Z	-1,298 ^a	-4,018 ^b	-6,284 ^b
Asymp. Sig. (2-tailed)	,194	,000	,000

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

Ranks		N	Mean Rank	Sum of Ranks
cr_MAE - ets_MAE	Negative Ranks	60 ^a	53,98	3239,00
	Positive Ranks	30 ^b	28,53	856,00
	Ties	0 ^c		
	Total	90		
cr_MAE - hw_MAE	Negative Ranks	66 ^d	51,00	3366,00
	Positive Ranks	24 ^e	30,38	729,00
	Ties	0 ^f		
	Total	90		
cr_MAE - arima_MAE	Negative Ranks	47 ^g	45,18	2123,50
	Positive Ranks	38 ^h	40,30	1531,50
	Ties	5 ⁱ		
	Total	90		

a. cr_MAE < ets_MAE

b. cr_MAE > ets_MAE

c. cr_MAE = ets_MAE

d. cr_MAE < hw_MAE

e. cr_MAE > hw_MAE

f. cr_MAE = hw_MAE

g. cr_MAE < arima_MAE

h. cr_MAE > arima_MAE

i. cr_MAE = arima_MAE

Test Statistics ^b			
	cr_MAE - ets_MAE	cr_MAE - hw_MAE	cr_MAE - arima_MAE
Z	-4,794 ^a	-5,305 ^a	-1,298 ^a
Asymp. Sig. (2-tailed)	,000	,000	,194

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
ets_MAE - cr_MAE	Negative Ranks	30 ^a	28,53	856,00
	Positive Ranks	60 ^b	53,98	3239,00
	Ties	0 ^c		
	Total	90		
ets_MAE - hw_MAE	Negative Ranks	43 ^d	35,43	1523,50
	Positive Ranks	47 ^e	54,71	2571,50
	Ties	0 ^f		
	Total	90		
ets_MAE - arima_MAE	Negative Ranks	34 ^g	30,85	1049,00
	Positive Ranks	56 ^h	54,39	3046,00
	Ties	0 ⁱ		
	Total	90		

a. ets_MAE < cr_MAE

b. ets_MAE > cr_MAE

c. ets_MAE = cr_MAE

d. ets_MAE < hw_MAE

e. ets_MAE > hw_MAE

f. ets_MAE = hw_MAE

g. ets_MAE < arima_MAE

h. ets_MAE > arima_MAE

i. ets_MAE = arima_MAE

Test Statistics^b

	ets_MAE - cr_MAE	ets_MAE - hw_MAE	ets_MAE - arima_MAE
Z	-4,794 ^a	-2,108 ^a	-4,018 ^a
Asymp. Sig. (2-tailed)	,000	,035	,000

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Ranks		N	Mean Rank	Sum of Ranks
hw_MAE - cr_MAE	Negative Ranks	24 ^a	30,38	729,00
	Positive Ranks	66 ^b	51,00	3366,00
	Ties	0 ^c		
	Total	90		
hw_MAE - ets_MAE	Negative Ranks	47 ^d	54,71	2571,50
	Positive Ranks	43 ^e	35,43	1523,50
	Ties	0 ^f		
	Total	90		
hw_MAE - arima_MAE	Negative Ranks	19 ^g	24,55	466,50
	Positive Ranks	70 ^h	50,55	3538,50
	Ties	1 ⁱ		
	Total	90		

a. hw_MAE < cr_MAE

b. hw_MAE > cr_MAE

c. hw_MAE = cr_MAE

d. hw_MAE < ets_MAE

e. hw_MAE > ets_MAE

f. hw_MAE = ets_MAE

g. hw_MAE < arima_MAE

h. hw_MAE > arima_MAE

i. hw_MAE = arima_MAE

Test Statistics ^c			
	hw_MAE - cr_MAE	hw_MAE - ets_MAE	hw_MAE - arima_MAE
Z	-5,305 ^a	-2,108 ^b	-6,284 ^a
Asymp. Sig. (2-tailed)	,000	,035	,000

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

Anexo G – Resultados do Teste Wilcoxon para erro percentual absoluto médio

Ranks		N	Mean Rank	Sum of Ranks
arima_MAPE - cr_MAPE	Negative Ranks	35 ^a	49,39	1728,50
	Positive Ranks	52 ^b	40,38	2099,50
	Ties	3 ^c		
	Total	90		
arima_MAPE - ets_MAPE	Negative Ranks	54 ^d	54,02	2917,00
	Positive Ranks	36 ^e	32,72	1178,00
	Ties	0 ^f		
	Total	90		
arima_MAPE - hw_MAPE	Negative Ranks	76 ^g	48,61	3694,00
	Positive Ranks	14 ^h	28,64	401,00
	Ties	0 ⁱ		
	Total	90		

a. arima_MAPE < cr_MAPE

b. arima_MAPE > cr_MAPE

c. arima_MAPE = cr_MAPE

d. arima_MAPE < ets_MAPE

e. arima_MAPE > ets_MAPE

f. arima_MAPE = ets_MAPE

g. arima_MAPE < hw_MAPE

h. arima_MAPE > hw_MAPE

i. arima_MAPE = hw_MAPE

Test Statistics ^c			
	arima_MAPE - cr_MAPE	arima_MAPE - ets_MAPE	arima_MAPE - hw_MAPE
Z	-,785 ^a	-3,499 ^b	-6,625 ^b
Asymp. Sig. (2-tailed)	,432	,000	,000

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

Ranks		N	Mean Rank	Sum of Ranks
cr_MAPE - ets_MAPE	Negative Ranks	56 ^a	52,63	2947,00
	Positive Ranks	34 ^b	33,76	1148,00
	Ties	0 ^c		
	Total	90		
cr_MAPE - hw_MAPE	Negative Ranks	68 ^d	49,04	3334,50
	Positive Ranks	22 ^e	34,57	760,50
	Ties	0 ^f		
	Total	90		
cr_MAPE - arima_MAPE	Negative Ranks	52 ^g	40,38	2099,50
	Positive Ranks	35 ^h	49,39	1728,50
	Ties	3 ⁱ		
	Total	90		

a. cr_MAPE < ets_MAPE

b. cr_MAPE > ets_MAPE

c. cr_MAPE = ets_MAPE

d. cr_MAPE < hw_MAPE

e. cr_MAPE > hw_MAPE

f. cr_MAPE = hw_MAPE

g. cr_MAPE < arima_MAPE

h. cr_MAPE > arima_MAPE

i. cr_MAPE = arima_MAPE

Test Statistics ^b			
	cr_MAPE - ets MAPE	cr_MAPE - hw MAPE	cr_MAPE - arima MAPE
Z	-3,619 ^a	-5,178 ^a	-,785 ^a
Asymp. Sig. (2-tailed)	,000	,000	,432

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

Ranks		N	Mean Rank	Sum of Ranks
ets_MAPE - cr_MAPE	Negative Ranks	34 ^a	33,76	1148,00
	Positive Ranks	56 ^b	52,63	2947,00
	Ties	0 ^c		
	Total	90		
ets_MAPE - hw_MAPE	Negative Ranks	43 ^d	43,72	1880,00
	Positive Ranks	47 ^e	47,13	2215,00
	Ties	0 ^f		
	Total	90		
ets_MAPE - arima_MAPE	Negative Ranks	36 ^g	32,72	1178,00
	Positive Ranks	54 ^h	54,02	2917,00
	Ties	0 ⁱ		
	Total	90		

a. ets_MAPE < cr_MAPE

b. ets_MAPE > cr_MAPE

c. ets_MAPE = cr_MAPE

d. ets_MAPE < hw_MAPE

e. ets_MAPE > hw_MAPE

f. ets_MAPE = hw_MAPE

g. ets_MAPE < arima_MAPE

h. ets_MAPE > arima_MAPE

i. ets_MAPE = arima_MAPE

Test Statistics ^b			
	ets_MAPE - cr_MAPE	ets_MAPE - hw_MAPE	ets_MAPE - arima_MAPE
Z	-3,619 ^a	-,674 ^a	-3,499 ^a
Asymp. Sig. (2-tailed)	,000	,500	,000

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

Ranks		N	Mean Rank	Sum of Ranks
hw_MAPE - cr_MAPE	Negative Ranks	22 ^a	34,57	760,50
	Positive Ranks	68 ^b	49,04	3334,50
	Ties	0 ^c		
	Total	90		
hw_MAPE - ets_MAPE	Negative Ranks	47 ^d	47,13	2215,00
	Positive Ranks	43 ^e	43,72	1880,00
	Ties	0 ^f		
	Total	90		
hw_MAPE - arima_MAPE	Negative Ranks	14 ^g	28,64	401,00
	Positive Ranks	76 ^h	48,61	3694,00
	Ties	0 ⁱ		
	Total	90		

a. hw_MAPE < cr_MAPE

b. hw_MAPE > cr_MAPE

c. hw_MAPE = cr_MAPE

d. hw_MAPE < ets_MAPE

e. hw_MAPE > ets_MAPE

f. hw_MAPE = ets_MAPE

g. hw_MAPE < arima_MAPE

h. hw_MAPE > arima_MAPE

i. hw_MAPE = arima_MAPE

Test Statistics ^c			
	hw_MAPE - cr_MAPE	hw_MAPE - ets_MAPE	hw_MAPE - arima_MAPE
Z	-5,178 ^a	-,674 ^b	-6,625 ^a
Asymp. Sig. (2-tailed)	,000	,500	,000

a. Based on negative ranks.

b. Based on positive ranks.

c. Wilcoxon Signed Ranks Test

Anexo H – Teste para a média das diferenças (amostras emparelhadas)

```
library(tstat)

Dados=read.csv('C:/Users/jpedropais/Desktop/erros_MAPE.csv', header =
TRUE, sep = ";")

CROSTON=Dados[,1]
ARIMA=Dados[,2]

t.test(ARIMA,CROSTON,mu=0,alternative="less",paired=TRUE)

summary(CROSTON)

summary(ARIMA)
```

Anexo I – Teste de Friedman

```
dados=scan('C:/Users/jpedropais/Desktop/Livro3.csv')

metodos=factor(c(rep(c("CR", "ETS", "HW", "ARIMA"), 90)))

refs=factor(c(rep("ref1", 4), rep("ref2", 4), rep("ref3", 4), rep("ref4", 4),
rep("ref5", 4), rep("ref6", 4), rep("ref7", 4), rep("ref8", 4), rep("ref9", 4),
rep("ref10", 4), rep("ref11", 4), rep("ref12", 4), rep("ref13", 4), rep("ref14",
4), rep("ref15", 4), rep("ref16", 4), rep("ref17", 4), rep("ref18", 4), rep("
ref19", 4), rep("ref20", 4), rep("ref21", 4), rep("ref22", 4), rep("ref23", 4),
rep("ref24", 4), rep("ref25", 4), rep("ref26", 4), rep("ref27", 4), rep("ref28",
4), rep("ref29", 4), rep("ref30", 4), rep("ref31", 4), rep("ref32", 4), rep("
ref33", 4), rep("ref34", 4), rep("ref35", 4), rep("ref36", 4), rep("ref37", 4),
rep("ref38", 4), rep("ref39", 4), rep("ref40", 4), rep("ref41", 4), rep("ref42",
4), rep("ref43", 4), rep("ref44", 4), rep("ref45", 4), rep("ref46", 4), rep("
ref47", 4), rep("ref48", 4), rep("ref49", 4), rep("ref50", 4), rep("ref51", 4),
rep("ref52", 4), rep("ref53", 4), rep("ref54", 4), rep("ref55", 4), rep("ref56",
4), rep("ref57", 4), rep("ref58", 4), rep("ref59", 4), rep("ref60", 4), rep("
ref61", 4), rep("ref62", 4), rep("ref63", 4), rep("ref64", 4), rep("ref65", 4),
rep("ref66", 4), rep("ref67", 4), rep("ref68", 4), rep("ref69", 4), rep("ref70",
4), rep("ref71", 4), rep("ref72", 4), rep("ref73", 4), rep("ref74", 4), rep("
ref75", 4), rep("ref76", 4), rep("ref77", 4), rep("ref78", 4), rep("ref79", 4),
rep("ref80", 4), rep("ref81", 4), rep("ref82", 4), rep("ref83", 4), rep("ref84",
4), rep("ref85", 4), rep("ref86", 4), rep("ref87", 4), rep("ref88", 4), rep("
ref89", 4), rep("ref90", 4)
))

friedman.test(dados,groups=metodos,blocks=refs)

require(PMCMR)

posthoc.friedman.conover.test(y=dados, p.adjust="none",
groups=metodos, blocks=refs)
```

Anexo J – Teste ANOVA

```
library(car)
dados=scan('C:/Users/jpedropais/Desktop/Livro3.csv')

metodos.n = c(1,2,3,4)
metodos.f = as.factor(metodos.n)
metodos.frame = data.frame(metodos.f)

metodos.model <- lm(dados~ 1)

analysys = Anova(metodos.model,idata=metodos.frame,idesign=~metodos.f)

summary(analysys)
```